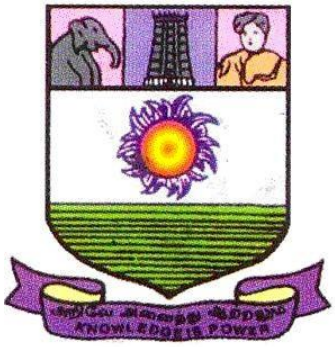
Web Technologies



DEPARTMENT Library and Information Science
Directorate of Distance and Continuing Education
Manonmaniam Sundaranar University
Tirunelveli – 627012

Course material prepared by

Dr R. Kavitha
Assistant Professor & Assistant Librarian
Mother Teresa Women's University
Kodaikanal,-624101, Tamil Nadu, India.

METHODS OF ASSESSMENT

Cognitive Level	Description
Remembering (K1)	The lowest level of questions require students to recall information from the course content. Knowledge questions usually require students to identify information in the textbook.
Understanding (K2)	Understanding of facts and ideas by comprehending organizing, comparing, translating, interpolating and interpreting in their own words. The questions go beyond simple recall and require students to combine together.
Application (K3)	Students have to solve problems by using/applying a concept learned in the classroom. Students must use their knowledge to determine an exact response.
Analyze (K4)	Analyzing the question is one of the tasks the students to breakdown something into its component parts. Analyzing requires students to identify reasons, causes or motives and reach conclusions or generalizations.
Evaluate (K5)	Evaluation requires an individual to make judgment on something. Questions to be asked to judge the value of an idea, a character, a work of art, or a solution to a problem. Students are engaged in decision-making and problem-solving. Evaluation questions do not have single right answers.
Create (K6)	The questions of this category challenge students to get engaged in creative and original thinking. Developing original ideas and problem-solving skills.

Pre-Requisites

Basic knowledge of computer fundamentals and internet usage. Familiarity with operating systems and networking concepts. Understanding of information retrieval and web browsing.

Learning Objectives

- To understand the evolution, growth, and architecture of the World Wide Web and Internet.
- To learn the principles of web design including markup languages (SGML, HTML, XML) and web browsers.
- To explore social media platforms and their applications in libraries.
- To gain knowledge about search engines, intranets, extranets, and search engine optimization.
- To understand web security concepts including threats, cyber crime, and security measures.

COURSE STRUCTURE

UNITS	Course Contents

UNIT I: WWW: An Overview	Evolution and growth of Web; Invisible Web; Internet Architecture: Hardware & Software Components, Client/Server Principle, Routers, Connection Types (Dial-up, ISDN, DSL, Cable, Satellite, Wireless), ISP, Protocols (TCP, IP, UDP, FTP, TFTP, Telnet, HTTP), Modem, Uniform Resource Locator (URL), IP Address (Public, Private, Dynamic, Static), Domain Name System (DNS), Web Servers (Apache).
UNIT II: Web Design Principles	Web Design Principles (Easy Navigation, Responsive Design, Color Scheme, UI, Content, Performance, Feedback); Markup Languages: SGML, HTML (Tags, Attributes, Data Types), XML (Well-formed, Valid, DTD, Schema, Parsers, DOM); Web Browsers: Internet Explorer, Mozilla Firefox, Google Chrome – History, Functions, Types, Architecture; Communication Tools: Email, Instant Messaging, VoIP, Video Conferencing.
UNIT III: Social Media and Web Development	Social Media – Definition, Business Applications, Benefits, Challenges, Types, Examples; Web Development Technologies: Browsers, HTML & CSS, Web Development Frameworks (Angular, Ruby on Rails, Yii, Meteor, Express.js, Zend, Django, Laravel), Programming Languages (JavaScript, Ruby, Elixir, Scala), Protocols, API, Data Formats (JSON, XML, CSV), Client and Server, Design in the Browser, Tools (Bootstrap, Style Guide, Chrome Developer Tools); Web Hosting and Publishing.
UNIT IV: Search Engines and Networks	Search Engines – Introduction, Market Share, Major Search Engines (Google, Yahoo, Bing, Ask), Directories; Search Provider Relationships; Components of Search Engine (Crawler, Index, Search Software); Ranking Factors (On-page, Off-page, Link Analysis, Link Popularity); Search Engine Spam (Techniques, Prevention); Intranet and Extranet – Components, Prerequisites, Services, Differences; Search Engine Components and Architecture; Design and Evaluation of Search Engines.
UNIT V: Web Security	Internet Security – Network Security Goals (Confidentiality, Integrity, Availability, Privacy, Authenticity); Security Engineering; Security Requirement Engineering; Viruses, Worms, Malware, Spyware, Adware, Trojans, Botnets; Defending Against Network Security Threats (Virus Scanners, IDS, Firewalls); Cyber Crime – Definition, Impact; Information Technology Act 2000 and Amendment Act 2008; Digital Signature and Electronic Signature; Electronic Governance; Firewalls, Antivirus, Anti-spyware.

TEXT BOOKS

S.No.	Book Details
1.	Duckett, Jon. (2014). *HTML and CSS: Design and Build Websites*. Wiley.
2.	Robbins, Jennifer Niederst. (2017). *Learning Web Design: A Beginner's Guide to HTML, CSS, JavaScript, and Web Graphics*. 4th Ed. O'Reilly Media.
3.	Morville, Peter & Rosenfeld, Louis. (2016). *Information Architecture for the World Wide Web: Designing Large-Scale Web Sites*. 3rd Ed. O'Reilly Media.

REFERENCE BOOKS

S.No.	Book Details
1.	Hartl, Michael & Prochazka, Aurelius. (2017). *RailsSpace: Building a Social Networking Website with Ruby on Rails*. Addison-Wesley Professional.
2.	Kalbach, James. (2017). *Designing Web Navigation: Optimizing the User Experience*. O'Reilly Media.
3.	Kurose, James F. & Ross, Keith W. (2017). *Computer Networking: A Top-Down Approach*. 7th Ed. Pearson.
4.	Stallings, William. (2017). *Network Security Essentials: Applications and Standards*. 6th Ed. Pearson.
5.	Tanenbaum, Andrew S. & Wetherall, David J. (2013). *Computer Networks*. 5th Ed. Pearson.

WEB SOURCES

S.No.	Web Source
1.	W3Schools Online Web Tutorials: https://www.w3schools.com
2.	Mozilla Developer Network (MDN): https://developer.mozilla.org
3.	World Wide Web Consortium (W3C): https://www.w3.org
4.	Internet Engineering Task Force (IETF): https://www.ietf.org
5.	Apache HTTP Server Project: https://httpd.apache.org

COURSE OUTCOMES

CO No.	Course Outcome	Cognitive Level
CO1	Remember and recall the fundamental concepts of WWW, Internet architecture, and web technologies.	K1
CO2	Understand and explain the principles of web design, markup languages, and communication tools.	K2
CO3	Apply knowledge of search engines, intranets, and extranets for effective information management.	K3
CO4	Analyze the features of different web development frameworks, programming languages, and social media platforms.	K4
CO5	Evaluate web security threats and apply appropriate security measures including legal provisions.	K5

K1-Remember; K2-Understand; K3-Apply; K4-Analyze; K5-Evaluate

UNIT I: WWW: An Overview

Content: Evolution and growth of Web; Invisible Web; Internet Architecture (Hardware & Software Components, Client/Server Principle, Routers, Connection Types, ISP, Protocols, Modem); URL; IP Address; Domain Name System; Web Servers (Apache).

UNIT-1

INTRODUCTION TO INTERNET

A worldwide computer network giving an assortment of data and correspondence facilities, comprising of interconnected organizations utilizing normalized correspondence conventions. The guide is also accessible on the Internet. The Internet is the worldwide arrangement of interconnected computer networks that utilization of the Internet protocol suite (TCP/IP) to interface gadgets around the world. It is a network of networks that comprises of private, public, business, academic and government organizations of nearby the worldwide scope, connected by a wide cluster of electronic, remote, and optical systems networking advancements. The Internet conveys a huge scope of data assets and administrations.

HISTORY OF INTERNET

This wonderful device has all set of experiences that hold its underlying foundations in the cold war situation. A need was acknowledged to associate the top colleges of the United States so they can share all the examination information without having an over the top delay. This endeavor was a consequence of Advanced Research Projects Agency (ARPA) which was framed toward the finish of 1950s, just after the Russians had climbed the space time with the send off of Sputnik. After the ARPA got achievement in 1969, it didn't take the specialists long to comprehend that how much potential would this interconnection be able to have the device. In 1971 Ray Tomlinson made a framework to send electronic mail. This was a major advance really taking shape as this opened entryways for remote computer getting to telnet.

During this time, thorough the paper work was being done in all the first class research organizations. From giving each and every computer a location to setting out the principles, everything was getting written down. 1973 saw the arrangements for the fundamental TCP/IP and Ethernet administrations. Toward the finish of 1970s, Usenet bunches had surfaced up. When the 80s had begun, IBM concocted its internet in view of Intel 8088 processor which was broadly involved by the students and colleges for it’s addressed the motivation behind simple registering. By 1982, the Defense Agencies made the

TCP/IP obligatory and the term internet was begat. The space name administrations showed

up in the year 1984 which is additionally the time around which different web based denoted their presentation. A worm or rust the internet's, assaulted in 1988 and incapacitated more than 10% of the internet frameworks from all over the world. While the greater part of the analysts viewed it has a chance to improve processing as it was in adolescent stage, a lot of internet organizations became keen on analyzing the centers of the malware which came about to the arrangement of Computer Emergency Rescue Team (CERT). Soon after the world moved past with the internet worm, World Wide Web appeared. Found by Tim Berners-Lee, World Wide Web was viewed as a support of interface reports in sites using hyperlinks.

WWW

The World Wide Web, abbreviated as WWW it is a data space where reports and other web assets are recognized by Uniform Resource Locators (URLs), interlinked by hypertext interfaces, and can be gotten through the Internet. English researcher TimBerners-Lee imagined the World Wide Web in 1989. He composed the primary internet browser computer program in 1990 while utilized at CERN in Switzerland. The Web program was delivered external CERN in 1991, first to other exploration organizations beginning in January 1991 and the overall population on the Internet in August 1991.

The World Wide Web has been integral to the improvement of the Information Age and is the essential device billions of individuals to use and connect on the Internet. Website pages are essentially message records designed and commented on with Hypertext Markup Language (HTML) and ordinarily known as the Web, It's an arrangement of interlinked hypertext reports accessed through the Internet. With a web program, one can see website pages that might contain text, pictures, recordings, and other interactive media and explore between them through hyperlinks.

Implanted hyperlinks license clients to explore between the web pages. Numerous pages with a typical subject, a typical space name or both make up a site. The site content can to a great extent be given by the distributer or intelligently where clients contribute content or the substance relies on the clients and their activities. Websites might be generally useful, fundamentally for amusement or to a great extent for business, legislative or non-administrative hierarchical purposes.

DEVELOPMENT AND GROWTH OF WEB

Between the summers of 1991 and 1994, the heap on the main Web server ("info.cern.ch") rose consistently by an element of 10 consistently. In 1992 scholarly community, and in 1993 industry, was paying heed. The WWW - Internet Consortium is framed in September 1994, with a base at MIT is the USA, INRIA in France, and presently also at Keio University in Japan. With the emotional surge of rich material of different types onto the Web during the 1990s, the first piece of the fantasy is to a great extent realized, albeit still not very few individuals in practice and by approaches to natural hypertext creation instruments. The second part presently can't seem to occur, yet there are signs and plans which make us certain. The extraordinary requirement for information about information to assist us with arranging, sort, pay for own data is driving the plan of dialects

for the web intended for handling by machines, instead of peopling. The trap of comprehensible record is being converged with a trap of machine-justifiable information. The capability of the combination of people and machines cooperating and imparting through the web could be immense.

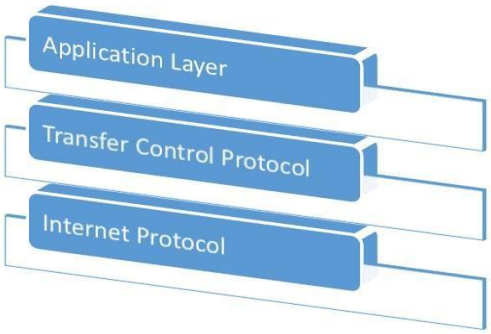
INVISIBLE WEB

Web indexes are could be sense, the heartbeat of the web; "Googling" has turned into a piece of ordinary discourse and is even perceived by Merriam-Webster as a linguistically right action word. It's a not unexpected misinterpretation, however that Googling is a search term it will uncover each site out of there that tends to your search. Commonplace web search tools like Google, Yahoo or Bing really access just a minuscule division. The destinations that customary pursuits yield are essential for what's known as the Surface Web, which are included listed of pages that a web search tool's web crawlers are customized to recover.

Most of the Internet lies in the invisible Web, at times alluded to as the Invisible Web. The real size of the invisible Web is difficult to quantify, however numerous specialists gauge it is multiple times the size of the web as far as we might be concerned. Seeing how surface pages are filed via web search tools can assist you with getting what's truly going on with the invisible Web. In early days, registering power and storage space was at such an exceptional that web crawlers recorded a negligible number of pages, frequently putting away just fractional substance. The technique behind looking through users goals; early Internet clients for the most part looked for research, so the primary web crawlers search engines straight forward inquiries that students or different analysts were probably going to make. Query items comprised of real substance that a web search tool had stored.

INTERNET ARCHITECTURE

Web browsing is a meta-network, which refers to an assemblage of thousands of unmistakable organizations interfacing with a typical convention. In basic terms, it is referred as an internet work that is associated utilizing protocols. In that convention utilized is TCP/IP. This protocol interfaces any two organizations that vary in hardware, software and design. TCP/IP gives start to finish transmission, i.e., every single hub on one organization can speak with some other hub on the network. That Internet architecture consists of three layers are.,



The three layers of internet

Internet Protocol

In order to impart, we want our information to be embodied as Internet Protocol (IP) packets. These IP packets traverse a number of hosts in a network through steering to arrive at the objective. Anyway IP doesn't uphold blunder recognition and mistake recuperation and it is unequipped for identifying loss of packets.

Transmission Control Protocol

TCP means "Transmission Control Protocol". It gives start to finish transmission of information, i.e., from source to objective. It is an exceptionally intricate protocol as it upholds recovery of lost packets.

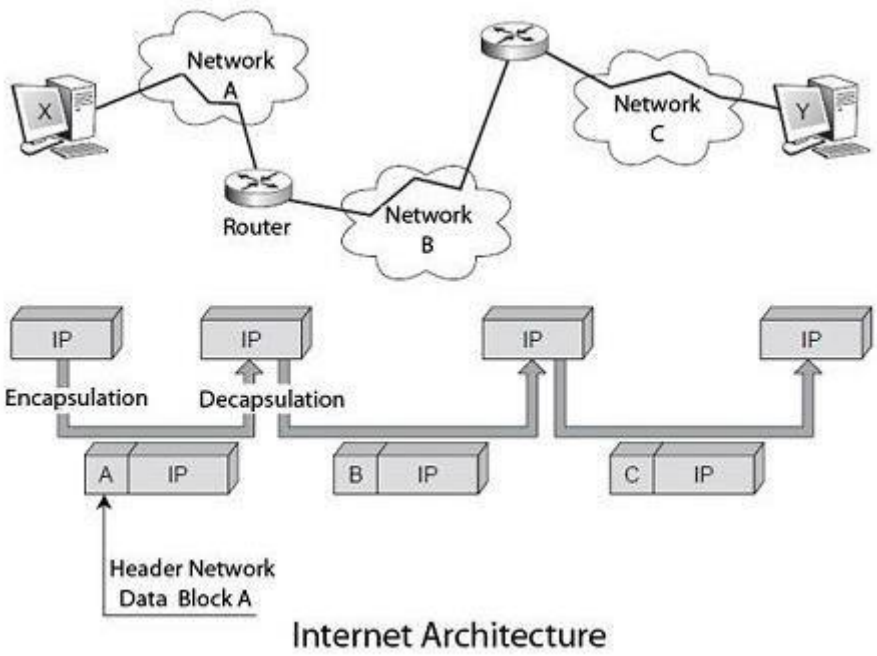
Application Protocol

Third layer is web architecture it’s the application layer which has various conventions on which the internet providers are constructed. A portion of the instances of internet providers incorporate email (SMTP works with email features), file transfer (FTP works with document move highlight), etc.

INTERNET ARCHITECTURE STRUCTURE

In the late 1960s, the US Department of Defense decides to make an extensive network from many small networks, all different, which begin to abound everywhere in North America. We had to find a way for these networks to coexist and give them outdoor visibility, the same for all users. Hence, Inter Network (interline), abbreviated as the Internet, data this network of networks.

The **Internet architecture** is based on a simple idea: ask all networks to carry a single packet type, a specific format, the IP protocol. Besides, this IP packet must have an address defined with sufficient generality to identify each computer and terminals scattered throughout the world. This architecture is illustrated in Figure.



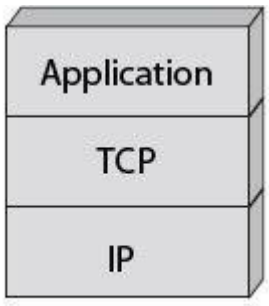
Internet Architecture

The user who wishes to make on this internetwork must store its data in IP packets delivered to the first network to cross. This first network encapsulates the IP packet in its packet structure, package A, which circulates in this form until an exit door, where it is decapsulated to retrieve the IP packet. The IP address is examined to locate, thanks to a routing algorithm, the following network to cross, and so on until arriving at the destination terminal.

To complete the IP, the US defense added the TCP protocol; specify the nature of the interface with the user. This protocol further determines how to transform a stream of bytes in an IP packet while ensuring the quality of transport of this IP packet. Both protocols, assembled under the TCP / IP abbreviation, are in the form of a layered architecture. They correspond to the packet level and message-level reference model.

The Internet model completed with a third layer called the application level, which includes different protocols for building Internet services. Email (SMTP), file transfer (FTP), the

transfer of hypermedia pages, transfer of distributed databases (World Wide Web), etc., are some of these services. The figure shows the three layers of Internet architecture.



The Three Layers of the Internet

IP packets are independent of each other and are individually routed in the network by interconnecting devices, subnets, routers. The quality of service offered by IP is minimal and provides no detection of lost or possibility of error recovery packages.

TCP combines the functionality of the message-level reference model. It is a fairly complex protocol with many options for solving all packet loss problems in the lower levels. In particular, a lost fragment can be recovered by retransmission on the stream of bytes. TCP uses a connection-oriented mode.

The flexibility of the Internet architecture can sometimes be a default. The extent that global optimization of the network is carried out by sub-network subnet, by a succession of local optimizations. It does not allow a homogeneous function in different subnets traversed. Another essential feature of this architecture is to place the entire control system, to say, intelligence and control of the network, in the terminal machine, leaving virtually nothing in the network, at least in the current version, IPv4, the IP protocol. The control intelligence is in the TCP software on the PC connected to the network.

It is the TCP protocol that takes care of sending more or fewer packets according to network load. Precise control window the maximum number of unacknowledged fragments that may be issued. The TCP window control increases or decreases the traffic following the time required to complete a round trip. Over this time increases, considering the more congested network, the transmission rate must decrease to counter saturation. In return, the infrastructure cost is meagre; no intelligence is not in the network. The service provided by

the network of networks corresponds to a quality called the best effort, which means that the network does its best to carry the traffic. In other words, the service quality is not assured.

The new generation of IP, IPv6, introduces new features that make the network nodes smarter. The new generation of routers comes with QoS management algorithms, which allow them to provide transportation that can meet time constraints or packet loss. We expect the arrival of IPv6 for ten years, but it's still IPv4 IP that governs the world. Because every new need is achievable with IPv6, IPv4 has been able to find the algorithms needed to do as well.

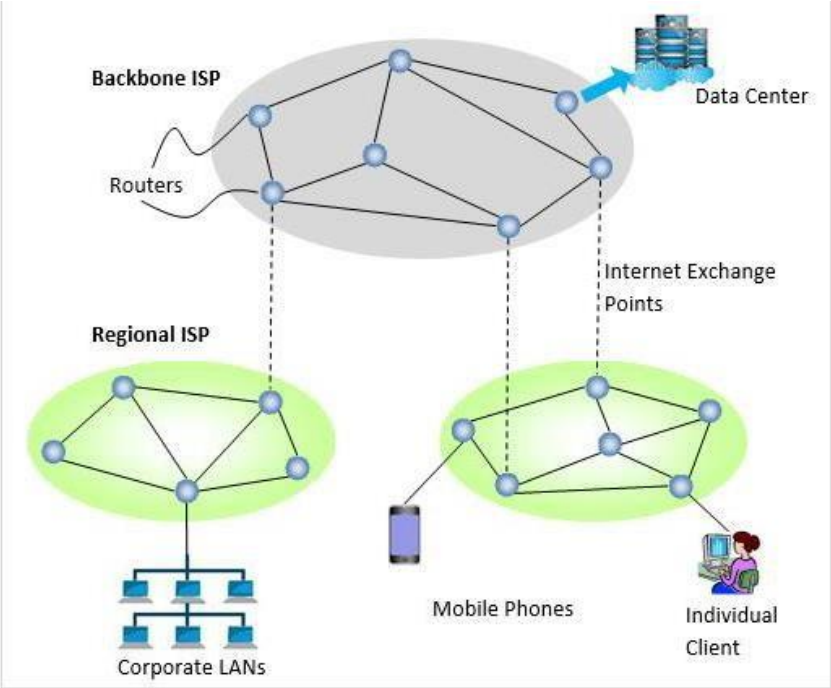
In IPv4, each new customer is treated the same way as those already connected, with resources being distributed equitably among all users. The resource allocation policies of telecom operator's networks are different since, on these networks, a customer who already has a certain quality of service does not suffer any penalty because of the arrival of a new customer. As discussed, the now advocated solution in the Internet environment is to encourage customers with real-time requirements, using appropriate protocols, using priority levels.

The IP protocol for thirty years but remained almost confidential for twenty years before taking off, unless its properties resulted from the failure of the protocols directly related to the reference model, too many and often incompatible. The IP world growth comes from the simplicity of its protocol, with very few options, and it's free.

The architecture of the Internet is ever-changing due to continuous changes in the technologies as well as the nature of the service provided. The heterogeneity and vastness of the Internet make it difficult to describe every aspect of its architecture. Internet architecture is a meta-network, which refers to a congregation of thousands of distinct networks interacting with a common protocol. In simple terms, it is referred as an internetwork that is connected using protocols. Protocol used is **TCP/IP**.

The overall architecture can be described in three levels −

1. Backbone ISP (Internet Service Provider)
2. Regional ISPs
3. Clients



Backbone ISP (Internet Service Provider) − Backbone ISPs are large international backbone networks. They are equipped with thousands of routers and store enormous amounts of information in data centers, connected through high bandwidth fiber optic links. Everyone needs to connect with a backbone ISP to access the entire Internet.

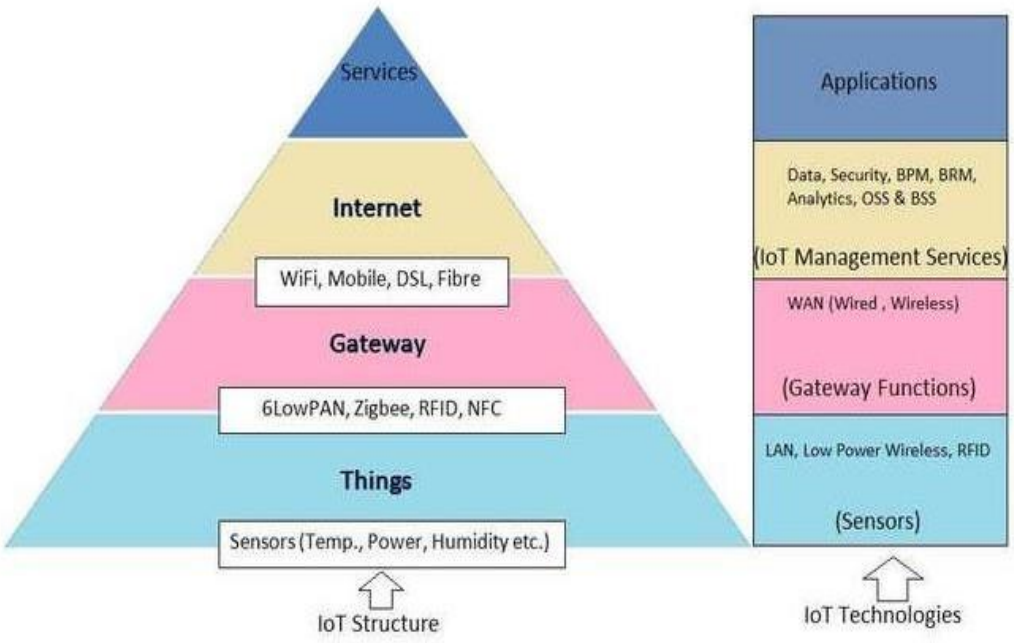
There are different ways through which a client can connect to the ISP. A commonly used way is DSL (Digital Subscriber Line) which reuses the telephone connection of the user for transmission of digital data. The user uses a dial-up connection instead of the telephone call. Connectivity is also done by sending signals over cable TV system that reuses unused cable TV channels for data transmission. For high-speed Internet access, the connectivity can be done through FTTH (Fiber to the Home), that uses optical fibers for transmitting data. Nowadays, most Internet access is done through the wireless connection to mobile phones from fixed subscribers, who transmit data within their coverage area.

INTERNET ARCHITECTURE HARDARE & SOFTWARE COMPONENTS

As we know IoT (Internet of Things) has been evolving at very rapid rate. Due to this, research on new IoT devices and IoT wireless technologies are also rising to bring the IoT products at cheaper and faster rate.
IoT has been classified into two categories viz. people to things referred as C2B (Customer to Business) and things to things or machine to machine referred as M2M.

People to Things involves IoT devices available at home such as wearables, fitness related devices, connected goods etc. M2M involves everything related to manufacturing and automation industry.

Let us understand basics of **IoT architecture**. As we know IoT system consists of three main parts viz. sensors, network connectivity and data storage applications. The same has been depicted in figure-1. As shown in the figure, Sensors in the IoT devices either communicate directly with the central server for data storage or communicate via gateway devices.



Sensors for various applications are used in different IoT devices as per different applications such as temperature, power, humidity, proximity, force etc.

Gateway takes care of various wireless standard interfaces and hence one gateway can handle multiple techologies and multiple sensors. The typical wireless technologies used widely are 6LoWPAN, Zigbee, Zwave, RFID, NFC etc. Gateway interfaces with cloud using backbone wireless or wired technologies such as WiFi, Mobile , DSL or Fibre.

As shown IoT supports both IPv4 and IPv6 protocols. Due to support of IPv6 which has about 128 bit long IP address length, there are enough addresses available to growing demand of IoT devices. DTN (Delay Tolerant Networks) is the unique feature of IoT which takes care

of large variable delay requirement of IoT based networks compare to traditional computer networks.

As shown , IoT service providers offer varied QoS with different pricing and design need for memory, CPU and battery consumption.

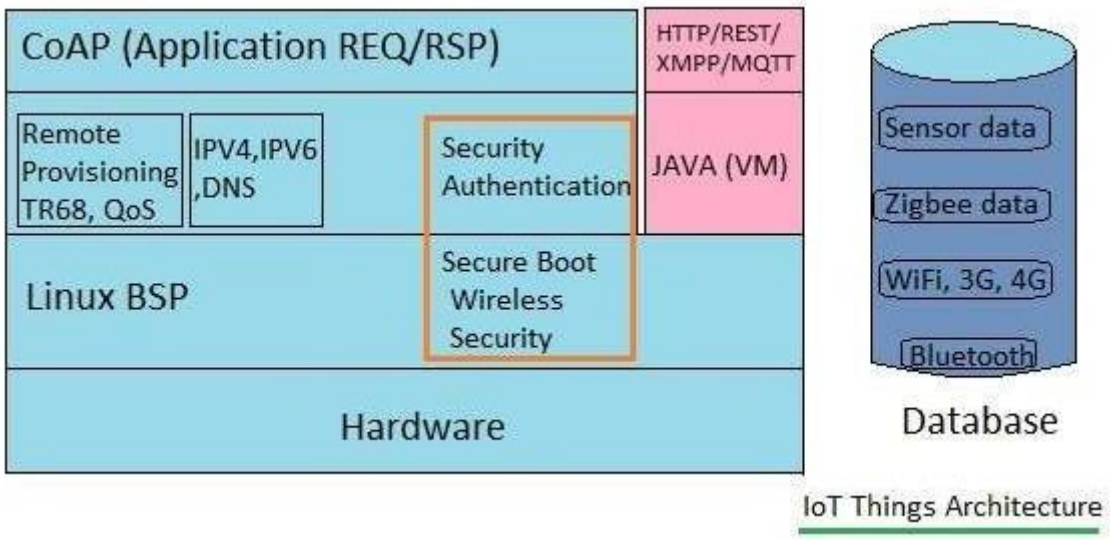
IoT Hardware Architecture

As we have seen that IoT device consists of upper protocol stack and physical and RF layers. The system can be constructed using MCU (Micro-controller Units). Choice of MCU depends on system on chip resources, power required and interfaces needed as per different sensors. Memory requirements of IoT hardware also needed to be carefully studied.

In order to finalize the IoT hardware architecture following aspects need to be collected. These parameters will help finalize ideal IoT hardware prototype as well as cost of the required IoT hardware components.

- Type of sensors/actuators
- Communication interface type
- Amount of data to be captured and transmitted
- Frequency of the data transportation

IoT Architecture for Software



Typically IoT software architecture is based on open source components. Figure-2 above depicts IoT architecture commonly in use for most of the systems. As shown Linux is widely used; as it need not require to wait to finalize the target hardware and software development can go in parallel.

Now-a-days many companies are working to provide ready to use IoT frameworks for various IoT specific applications. CoAP protocol is used which is unique to IoT applications and offer common mechanism to communicate with IoT devices.

Vendors of IoT frameworks

Following table mentions IoT frameworks vendors, providers or developers.

Vendors, developers or providers of IoT FrameWorks
Thread Framework from Thread Group based on Zigbee and 6LoWPAN standards. This frameworks require gateway in order to communicate with the server.
Eclipse Open IoT
Open Interconnect Consortium known as IoTivity
Linear technology
Microsoft
Oracle

We have gone through, IoT architecture basics including IoT hardware architecture, IoT software architecture and IoT framework vendors.

Cloud Computing Architecture

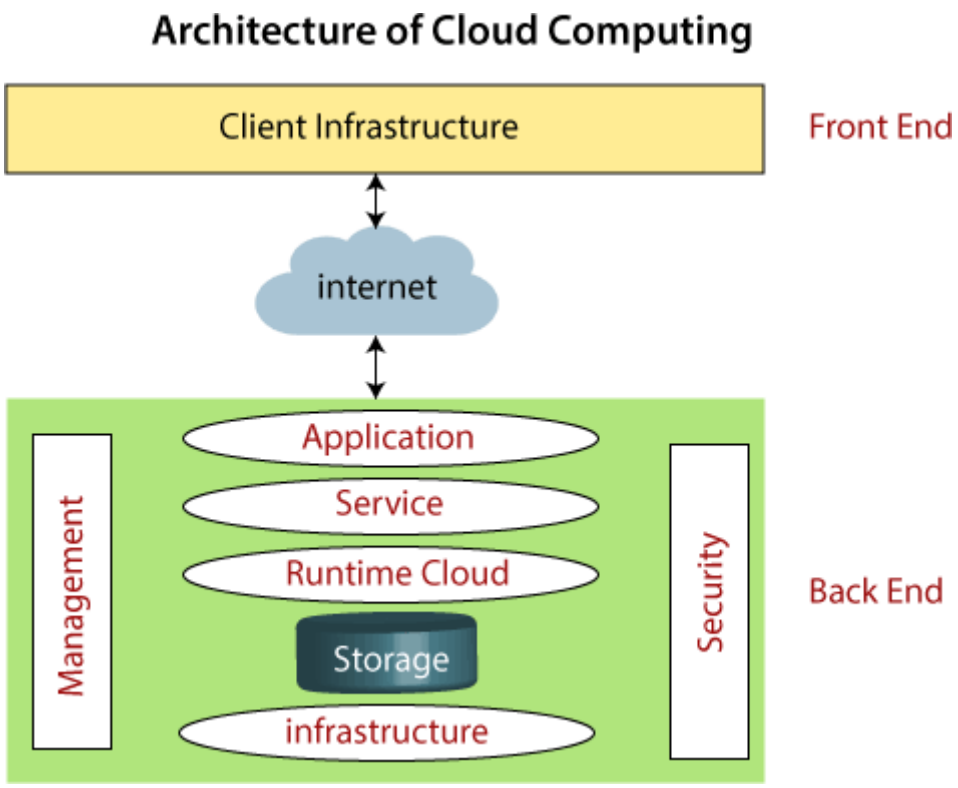
As we know, cloud computing technology is used by both small and large organizations to **store the information** in cloud and **access** it from anywhere at anytime using the internet connection.

Cloud computing architecture is a combination of **service-oriented architecture** and **event-driven architecture**.

Cloud computing architecture is divided into the following two parts -

- Front End
- Back End

The below diagram shows the architecture of cloud computing -



Front End

The front end is used by the client. It contains client-side interfaces and applications that are required to access the cloud computing platforms. The front end includes web servers (including Chrome, Firefox, internet explorer, etc.), thin & fat clients, tablets, and mobile devices.

Back End

The back end is used by the service provider. It manages all the resources that are required to provide cloud computing services. It includes a huge amount of data storage, security mechanism, virtual machines, deploying models, servers, traffic control mechanisms, etc.

Components of Cloud Computing Architecture

There are the following components of cloud computing architecture -

1. Client Infrastructure

Client Infrastructure is a Front end component. It provides GUI (Graphical User Interface) to interact with the cloud.

2. Application

The application may be any software or platform that a client wants to access.

3. Service

A Cloud Services manages that which type of service you access according to the client's requirement.

Cloud computing offers the following three type of services:

i. Software as a Service (SaaS) – It is also known as **cloud application services**. Mostly, SaaS applications run directly through the web browser means we do not require to download and install these applications. Some important example of SaaS is given below –

Example: Google Apps, Salesforce Dropbox, Slack, Hubspot, Cisco WebEx.

ii. Platform as a Service (PaaS) – It is also known as **cloud platform services**. It is quite similar to SaaS, but the difference is that PaaS provides a platform for software creation, but using SaaS, we can access software over the internet without the need of any platform.

Example: Windows Azure, Force.com, Magento Commerce Cloud, OpenShift.

iii. Infrastructure as a Service (IaaS) – It is also known as **cloud infrastructure services**. It is responsible for managing applications data, middleware, and runtime environments.

Example: Amazon Web Services (AWS) EC2, Google Compute Engine (GCE), Cisco Metapod.

4. Runtime Cloud

Runtime Cloud provides the **execution and runtime environment** to the virtual machines.

5. Storage

Storage is one of the most important components of cloud computing. It provides a huge amount of storage capacity in the cloud to store and manage data.

6. Infrastructure

It provides services on the **host level**, **application level**, and **network level**. Cloud infrastructure includes hardware and software components such as servers, storage, network devices, virtualization software, and other storage resources that are needed to support the cloud computing model.

7. Management

Management is used to manage components such as application, service, runtime cloud, storage, infrastructure, and other security issues in the backend and establish coordination between them.

8. Security

Security is an in-built back end component of cloud computing. It implements a security mechanism in the back end.

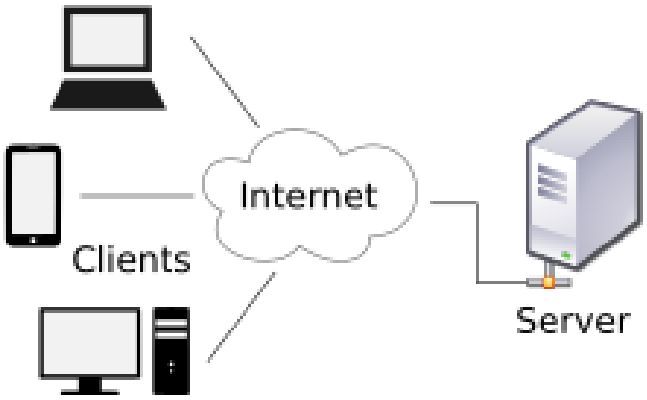
9. Internet

The Internet is medium through which front end and back end can interact and communicate with each other.

THE CLIENT/SERVER PRINCIPLE:

The client/server principle refers to the two components of a centralized computer network: client and server machines. Clients request information and servers send them the requested information. For example, when an individual uses his computer to look at a Web page, his computer acts as the client, and the computer hosting the Web page is the server. Browsers enable the connection between clients and servers. During the 1990s, Netscape, an outgrowth of the early browser Mosaic, and Internet Explorer were the dominant browsers. Eventually, Microsoft's Internet Explorer became the dominant browser. Other frequently used browsers are Mozilla Firefox and Safari.

Client-server model is a distributed application structure that partitions tasks or workloads between the providers of a resource or service, called servers, and service requesters, called clients. Often clients and servers communicate over a computer network on separate hardware, but both client and server may reside in the same system. A server host runs one or more server programs, which share their resources with clients. A client usually does not share any of its resources, but it requests content or service from a server. Clients, therefore, initiate communication sessions with servers, which await incoming requests. Examples of computer applications that use the client-server model are email, network printing, and the World Wide Web.



The Advantages of a Client-Server Network

The biggest advantage to using this setup is central management of the server. Only one server is used to host the resources that all the clients request and use. This is especially good for server administrators, because they only have to be in one place and can solve all the problems in one place, as well. Having to manually update several hundred servers would take much more time. One centrally managed server is the key to ease of management, and it is cost effective, too.

Another advantage of using one physical server is that the configuration is simple to set up and takes less time to troubleshoot. For instance, if there were a site with multiple servers providing redundant services, and it was having issues, it could take an extreme amount of work to effectively troubleshoot why services are being hindered. In a single server role, all troubleshooting takes place at one physical server, so it takes much less time.

WHAT IS A ROUTER AND HOW DOES IT WORK:

A router is a physical or virtual appliance that passes information between two or more packet-switched computer networks. A router inspects a given data packet's destination Internet Protocol address (IP address), calculates the best way for it to reach its destination and then forwards it accordingly.

A router is a common type of gateway. It is positioned where two or more networks meet at each point of presence on the internet. Hundreds of routers might forward a single packet as it moves from one network to the next on the way to its final destination. In the Open Systems Interconnection (OSI) model, routers are associated with the network layer.

Traditional routers are stand-alone devices that use proprietary software. In contrast, a virtual router is a software instance that performs the same functions as a physical router. Virtual routers typically run on commodity servers, either alone or packaged with other virtual network functions, like firewall packet filtering, load balancing and wide area network (WAN) optimization capabilities.

Other network devices, such as wireless access points and switches may include built-in router functionality. A router is a device that communicates between the internet and the devices in your home that connect to the internet. As its name implies, it “routes” traffic between the devices and the internet.

With the right kind of router in your home, you may be able to enjoy faster internet service, help protect your family from cyberthreats, and avoid those maddening Wi-Fi dead spots. You don’t have to be a computer genius to know what a good router has to offer. All it takes is to know what you need it for. Understanding how routers work will help you choose the right equipment for your home.

How do routers work:

A typical home has a range of internet-connected devices — personal computers, tablets, smartphones, printers, thermostats, smart TVs, and more. With your router, these devices form a network. A router directs incoming and outgoing internet traffic on that network in the fastest and most efficient way.

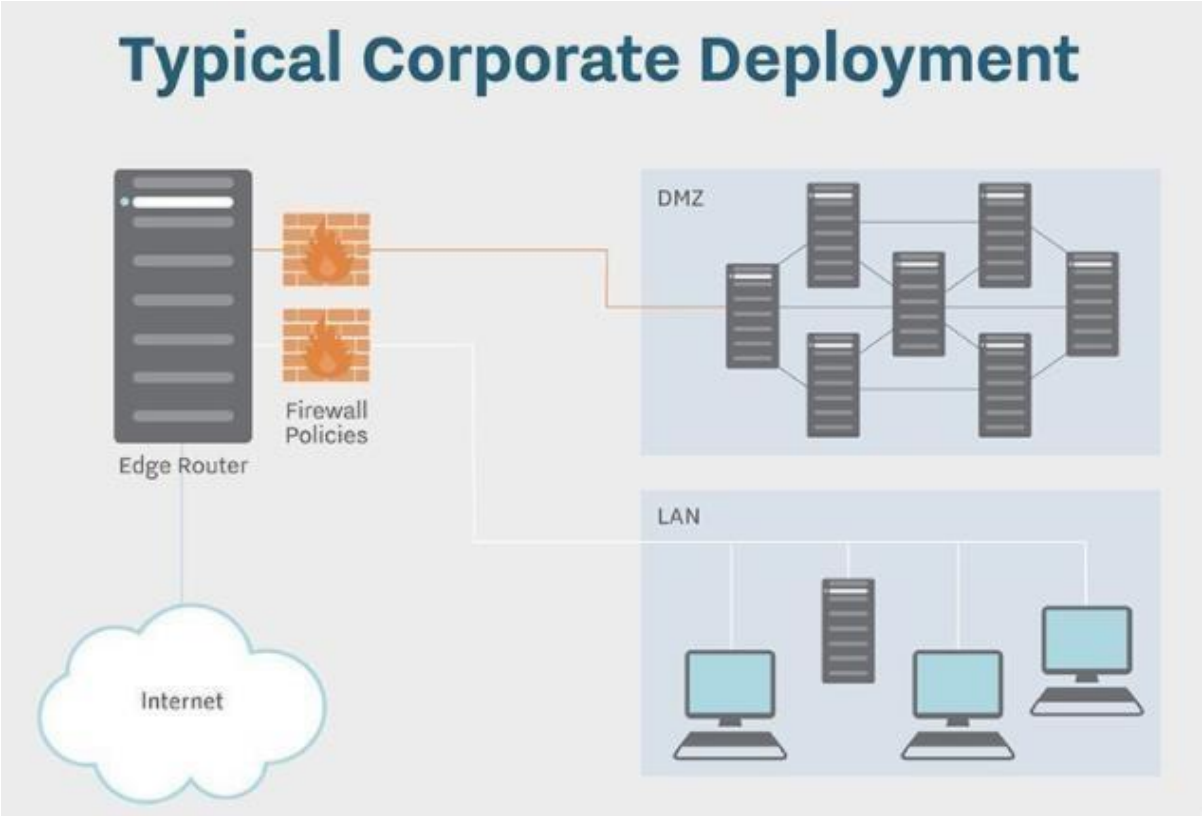
The information traveling on your home network could be an email, a movie, or a live feed from your baby cam, each of which takes up varying amounts of bandwidth. Making sure that information is delivered quickly and correctly is a big task — and getting bigger. As you add more and more devices — think Internet of Things — you ask your router to do more.

A router examines a packet header's destination IP address and compares it against a routing table to determine the packet's best next hop. Routing tables list directions for forwarding data to particular network destinations, sometimes in the context of other variables, like cost. They amount to an algorithmic set of rules that calculate the best way to transmit traffic toward any given IP address.

A routing table often specifies a default route, which the router uses whenever it fails to find a better forwarding option for a given packet. For example, the typical home office router directs all outbound traffic along a single default route to its internet service provider (ISP).

Routing tables can be static -- i.e., manually configured -- or dynamic. Dynamic routers automatically update their routing tables based on network activity, exchanging information with other devices via routing protocols.

Many routers also perform network address translation (NAT), shielding the private IP addresses of a local area network (LAN) by readdressing all outgoing traffic with a single shared public IP address. NAT helps both conserve globally valid IP addresses and improve network security.



How modems differ from routers

A router and your devices aren't the only components on your home network. There's also the modem. In fact, without the modem, all you'd have is your local network with no access to the internet.

The modem's job is to bring the internet service from your provider into your home. It then connects to your router, delivering that internet connectivity to your home network.

When most internet service was delivered over telephone lines, modems enabled communication between the digital devices in your home and the analog signals used on telephone lines. With today's internet connections, including cable and satellite, modems play a similar but different role.

What are the different types of routers?

When it comes to routers, there are only two types you'll need to consider:

1. **Wireless routers**. A wireless router connects directly to a modem by a cable. This allows it to receive information from — and transmit information to — the internet. The router then creates and communicates with your home Wi-Fi network using built-in antennas. As a result, all of the devices on your home network have internet access.
2. **Wired routers**. A wired router connects directly to computers through wired connections. They usually have a port that connects to the modem to communicate with the internet. Another port — or ports — allows the wired router to connect to computers and other devices to distribute information.
3. **Edge Router.** It seats at the edge of the backbone of the network and can connect to the core routers. It can be wired or wireless and will distribute internet data packets between one or more networks. But it will not distribute internet data packets within networks.
4. **Core Router.** It is designed to operate in the internet backbone or core. It supports multiple telecommunication interfaces of the highest speed and usage in the core internet. It can forward IP packets at full speed on all of them. It supports the routing protocol that is used in the core. It will distribute internet data packets within the network. But core will not distribute internet data packets between networks.
5. **Virtual Router.**

It is default for a computer sharing network. It functions as per the virtual router redundancy protocol (VRRP), it becomes active when the main or primary router fails or becomes disabled. It takes multiple routers in a group so that they can share a virtual IP address. It has a master for each group that handles IP packets. If the master fails while forwarding packets then other routers will take a position.

What to look for in a router

Most internet service providers (ISPs) give you a router and a modem — or a combination of the two — for a subscription fee that can add up over time. These routers may not be the best fit for your usage, so you might consider purchasing one that better fits your needs. Before buying a router, here are a few things to look for.

Wi-Fi coverage

Wi-Fi signals within a home largely depend on the size of the home and the barriers that prevent signals from reaching their destinations. Fireplaces, mirrors, and thick walls are just a few common obstacles that block Wi-Fi signals. Look for a router that has the capability to reach the far corners of your home. Also, look for one that has a mesh network to extend the Wi-Fi capabilities across the home.

Wi-Fi performance

Router technology has changed over time. Make sure you have a router that uses the latest technology and has updated firmware. MU-MIMO is one such new technology. It stands for multi-user, multiple-input, multiple-output technology. It allows Wi-Fi routers to communicate with multiple devices simultaneously. This decreases the wait time and improves network speed.

Wi-Fi security

Cybercriminals can penetrate your home network and install malware and viruses in your devices. They work with an arsenal of tools to gain access to your personal and financial information. Having a router that provides network level protection could help protect against cyberattacks at the port of entry. Look for a router that has built-in security features, like automatic updates, device quarantine, and signed firmware updates.

Wi-Fi controls

Routers have become a very important part of the connected home. Make sure you buy a router that you can control easily. The latest routers are easy to install and use. Some come with user-friendly apps that help you with guest networks, parental controls, user time limits, and network management.

Whether you are setting up a new router in your home or upgrading an existing one, make sure you get to know all the workings of your new router and if it is designed to meet your needs.

Router protocols

Routing protocols determine how a router identifies other routers on the network, keeps track of all possible destinations and makes dynamic decisions for where to send each network message. Popular protocols include:

Open Shortest Path First (OSPF) -- used to find the best path for packets as they pass through a set of connected networks. OSPF is designated by the Internet Engineering Task Force (IETF) as one of several Interior Gateway Protocols (IGPs).

Border Gateway Protocol (BGP) -- manages how packets are routed across the internet through the exchange of information between edge routers. BGP offers network stability that guarantees routers can quickly adapt to send packets through another reconnection if one internet path goes down.

Interior Gateway Routing Protocol (IGRP) -- determines how routing information between gateways will be exchanged within an autonomous network. The routing information can then be used by other network protocols to specify how transmissions should be routed.

Enhanced Interior Gateway Routing Protocol (EIGRP) -- evolved from IGRP. If a router can't find a route to a destination in one of these tables, it queries its neighbors for a route and they in turn query their neighbors until a route is found. When a routing table entry changes in one of the routers, it notifies its neighbors of the change instead of sending the entire table.

Exterior Gateway Protocol (EGP) -- determines how routing information between two neighbor gateway hosts, each with its own router, is exchanged. EGP is commonly used between hosts on the Internet to exchange routing table information.

Routing Information Protocol (RIP) -- the original protocol for defining how routers should share information when moving traffic among an interconnected group of local area networks. The largest number of hops allowed for RIP is 15, which limits the size of networks that RIP can support.

CONNECTION TYPES

There exist several ways to connect to the internet. Following are these connection types available:

1. Dial-up Connection
2. ISDN
3. DSL

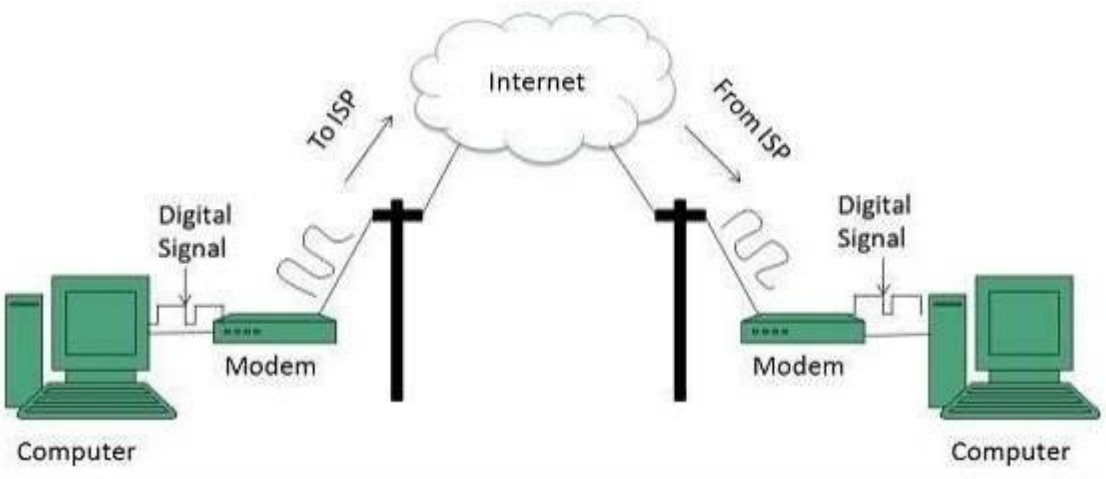
4. Cable TV Internet connections

5. Satellite Internet connections

6. Wireless Internet Connections

Dial-up Connection Dial-up connection uses telephone line to connect PC to the internet. It requires a modem to setup dial-up connection. This modem works as an interface between PC and the telephone line. There is also a communication program that instructs the modem to make a call to specific number provided by an ISP. Dial-up connection uses either of the following protocols:

1. Serial Line Internet Protocol (SLIP)
2. Point to Point Protocol (PPP)



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:

1. Basic Rate Interface (BRI)

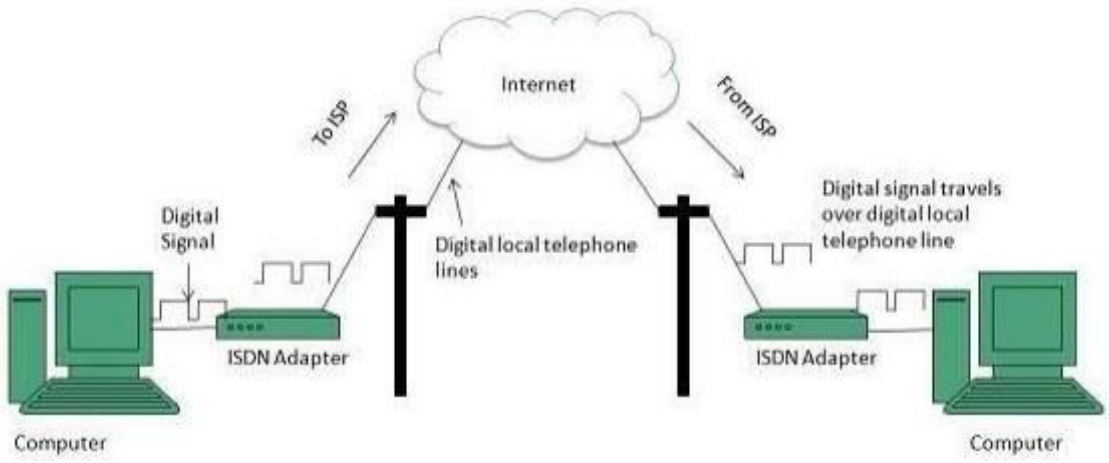
2. Primary Rate Interface (PRI)

Key points:

- The BRI ISDN consists of three distinct channels on a single ISDN line: t1o 64kbps B (Bearer) channel and one 16kbps D (Delta or Data) channels.

- The PRI ISDN consists of 23 B channels and one D channels with both have operating capacity of 64kbps individually making a total transmission rate of 1.54Mbps.

The following diagram shows accessing internet using ISDN connection:



DSL

DSL is acronym of Digital Subscriber Line.

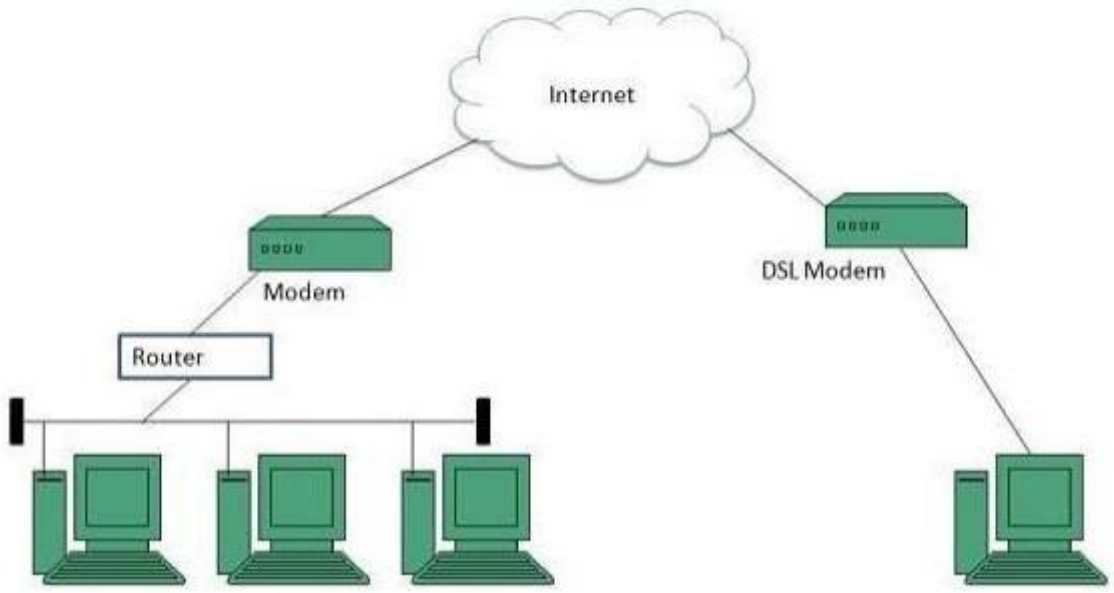
It is a form of broadband connection as it provides connection over ordinary telephone lines.

Following are the several versions of DSL technique available today:

1. **Asymmetric DSL (ADSL)**
2. **Symmetric DSL (SDSL)**
3. **High bit-rate DSL (HDSL)**
4. **Rate adaptive DSL (RDSL)**
5. **Very high bit-rate DSL (VDSL)**
6. **ISDN DSL (IDSL)**

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

The following diagram shows that how we can connect to internet using DSL technology:



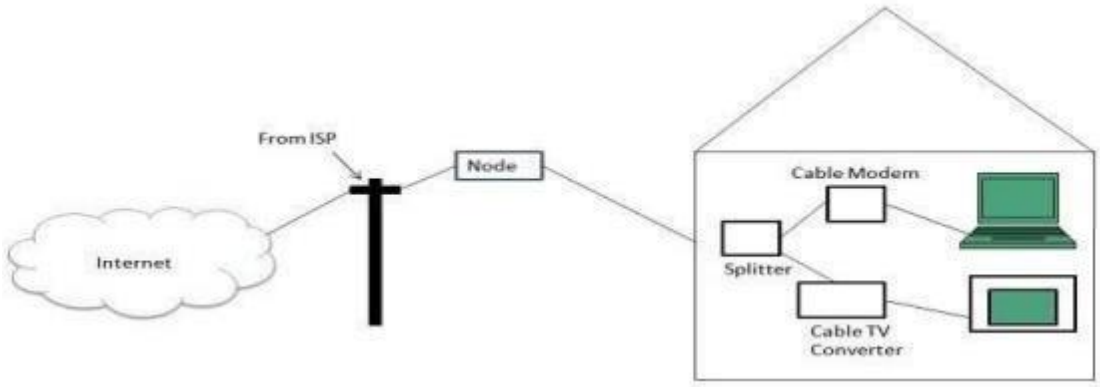
Cable TV Internet Connection

Cable TV Internet connection is provided through Cable TV lines. It uses coaxial cable which is capable of transferring data at much higher speed than common telephone line.

Key Points:

- A cable modem is used to access this service, provided by the cable operator.
- The Cable modem comprises of two connections: one for internet service and other for Cable TV signals.
- Since Cable TV internet connections share a set amount of bandwidth with a group of customers, therefore, data transfer rate also depends on number of customers using the internet at the same time.

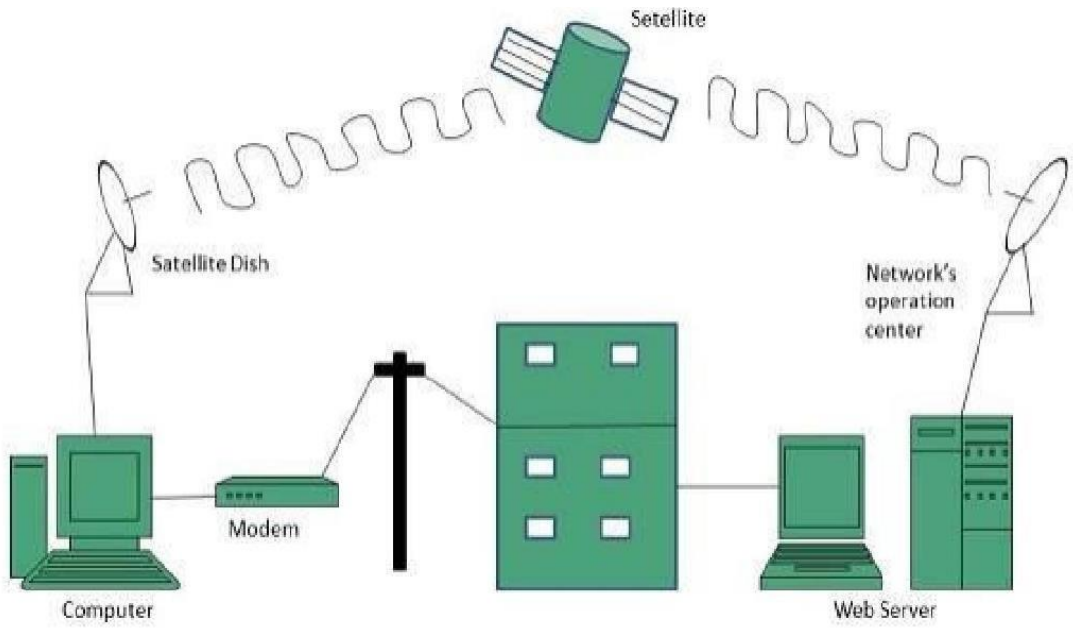
The following diagram shows that how internet is accessed using Cable TV connection:



Satellite Internet Connection

Satellite Internet connection offers high speed connection to the internet. There are two types of satellite internet connection: one way connection or two way connection. In one way connection, we can only download data but if we want to upload, we need a dialup access through ISP over telephone line. In two way connection, we can download and upload the data by the satellite. It does not require any dialup connection.

The following diagram shows how internet is accessed using satellite internet connection:



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.

Key Points:

- Wi Fi wireless technology is based on IEEE 802.11 standards which allow the electronic device to connect to the internet.

- Bluetooth wireless technology makes use of short-wavelength radio waves and helps to create personal area network (PAN).

Internet Protocols

Transmission Control Protocol (TCP)

TCP is a connection oriented protocol and offers end-to-end packet delivery.
It acts as back bone for connection.It exhibits the following key features:

- Transmission Control Protocol (TCP) corresponds to the Transport Layer of OSI Model.
- TCP is a reliable and connection oriented protocol.
- TCP offers:
    - Stream Data Transfer.
    - Reliability.
    - Efficient Flow Control
    - Full-duplex operation.
    - Multiplexing.
- TCP offers connection oriented end-to-end packet delivery.
- TCP ensures reliability by sequencing bytes with a forwarding acknowledgement number that indicates to the destination the next byte the source expect to receive.

It retransmits the bytes not acknowledged with in specified time period.

TCP Services

TCP offers following services to the processes at the application layer:

- Stream Delivery Service
- Sending and Receiving Buffers
- Bytes and Segments
- Full Duplex Service
- Connection Oriented Service
- Reliable Service

Stream Deliver Service

TCP protocol is stream oriented because it allows the sending process to send data as stream of bytes and the receiving process to obtain data as stream of bytes.

Sending and Receiving Buffers

It may not be possible for sending and receiving process to produce and obtain data at same speed, therefore, TCP needs buffers for storage at sending and receiving ends.

Bytes and Segments

The Transmission Control Protocol (TCP), at transport layer groups the bytes into a packet. This packet is called segment. Before transmission of these packets, these segments are encapsulated into an IP datagram.

Full Duplex Service

Transmitting the data in duplex mode means flow of data in both the

directions at the same time.

Connection Oriented Service

TCP offers connection oriented service in the following manner:

1. TCP of process-1 informs TCP of process – 2 and gets its approval.
2. TCP of process – 1 and TCP of process – 2 and exchange data in both the two directions.
1. After completing the data exchange, when buffers on both sides are empty, the two TCP's destroy their buffers.

Reliable Service

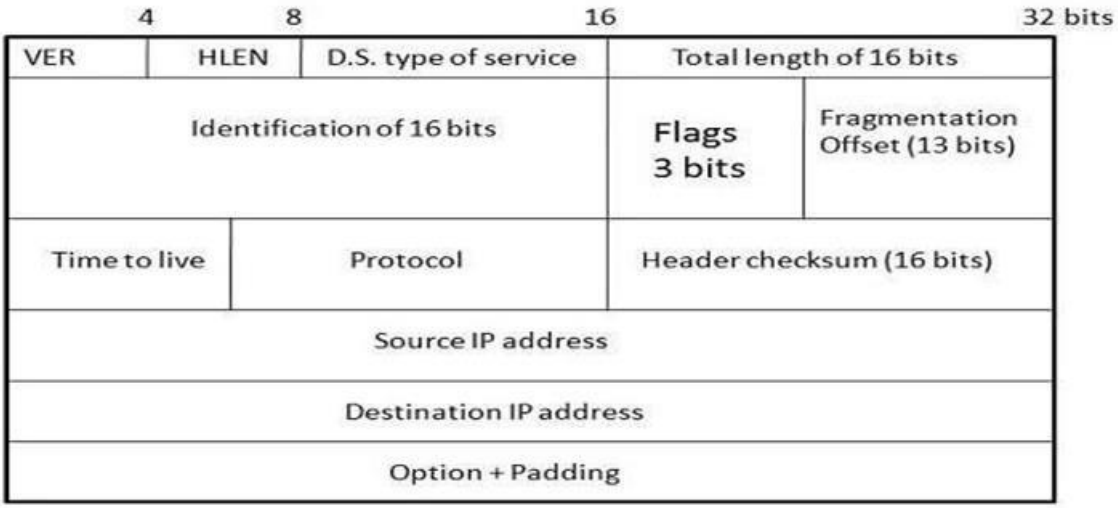
For sake of reliability, TCP uses acknowledgement mechanism.

Internet Protocol (IP)

Internet Protocol is **connectionless** and **unreliable** protocol. It ensures no guarantee of successfully transmission of data.

In order to make it reliable, it must be paired with reliable protocol such as TCP at the transport layer.

Internet protocol transmits the data in form of a datagram as shown in the following diagram:



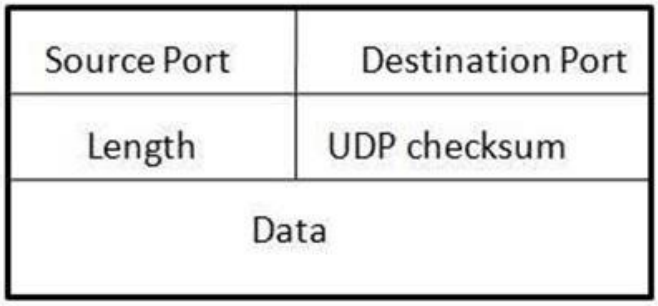
Points to remember:

- The length of datagram is variable.
- The Datagram is divided into two parts: **header** and **data.**
- The length of header is 20 to 60 bytes.
- The header contains information for routing and delivery of the packet.

User Datagram Protocol (UDP)

Like IP, UDP is connectionless and unreliable protocol. It doesn't require making a connection with the host to exchange data. Since UDP is unreliable protocol, there is no mechanism for ensuring that data sent is received.

UDP transmits the data in form of a datagram. The UDP datagram consists of five parts as shown in the following diagram:



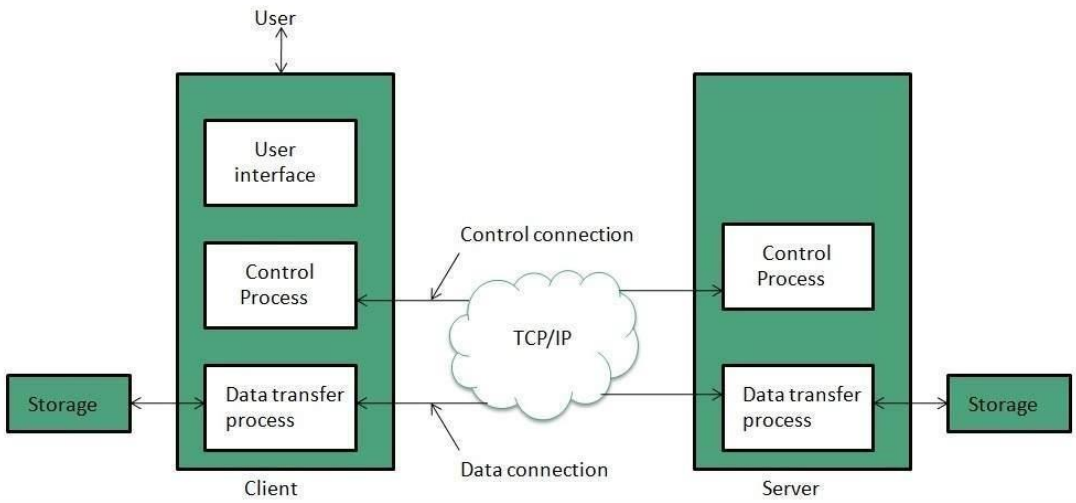
Points to remember:

- UDP is used by the application that typically transmit small amount of data at one time.
- UDP provides protocol port used i.e. UDP message contains both source and destination port number, that makes it possible for UDP software at the destination to deliver the message to correct application program.

File Transfer Protocol (FTP)

FTP is used to copy files from one host to another. FTP offers the mechanism for the same in following manner:

- FTP creates two processes such as Control Process and Data Transfer Process at both ends i.e. at client as well as at server.
- FTP establishes two different connections: one is for data transfer and other is for control information.
- **Control connection** is made between **control processes** while **Data Connection** is made between<="" b="">
- FTP uses **port 21** for the control connection and **Port 20** for the data connection.



Trivial File Transfer Protocol (TFTP)

Trivial File Transfer Protocol is also used to transfer the files but it transfers the files without authentication. Unlike FTP, TFTP does not separate control and data information. Since there is no authentication exists, TFTP lacks in security features therefore it is not recommended to use TFTP.

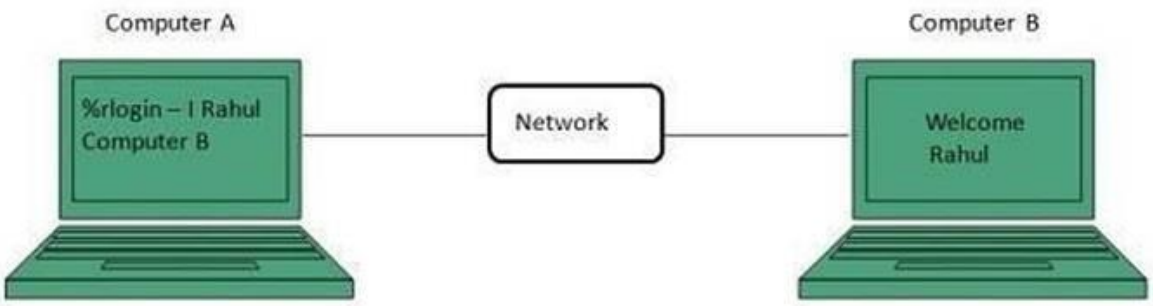
Key points

- TFTP makes use of UDP for data transport. Each TFTP message is carried in separate UDP datagram.
- The first two bytes of a TFTP message specify the type of message.
- The TFTP session is initiated when a TFTP client sends a request to upload or download a file.
- The request is sent from an ephemeral UDP port to the **UDP port 69** of an TFTP server.

S.No	Parameter	FTP	TFTP
1	Operation	Transferring Files	Transferring Files
2	Authentication	Yes	No
3	Protocol	TCP	UDP
4	Ports	21 – Control, 20 – Data	Port 3214, 69, 4012
5	Control and Data	Separated	Separated
6	Data Transfer	Reliable	Unreliable

Telnet

Telnet is a protocol used to log in to remote computer on the internet. There are a number of Telnet clients having user friendly user interface. The following diagram shows a person is logged in to computer A, and from there, he remote logged into computer B.



Hyper Text Transfer Protocol (HTTP)

HTTP is a communication protocol. It defines mechanism for communication between browser and the web server. It is also called request and response protocol because the communication between browser and server takes place in request and response pairs.

HTTP Request

HTTP request comprises of lines which contains:

- Request line
- Header Fields
- Message body

Key Points

- The first line i.e. the **Request line** specifies the request method i.e. **Get** or **Post.**
- The second line specifies the header which indicates the domain name of the server from where index.htm is retrieved.

HTTP Response

Like HTTP request, HTTP response also has certain structure. HTTP response contains:

- Status line

- Headers
- Message body

MODEM

Modem is a device that enables a computer to send or receive data over telephone or cable lines. The data stored on the computer is digital whereas a telephone line or cable wire can transmit only analog data.

Digital Data Waveform

Analog Data Waveform

The main function of the modem is to convert digital signal into analog and vice versa. Modem is a combination of two devices − **modulator** and **demodulator**. The **modulator** converts digital data into analog data when the data is being sent by the computer. The **demodulator** converts analog data signals into digital data when it is being received by the computer.

Types of Modem

Modem can be categorized in several ways like direction in which it can transmit data, type of connection to the transmission line, transmission mode, etc.

Depending on direction of data transmission, modem can be of these types −

- **Simplex** − A simplex modem can transfer data in only one direction, from digital device to network (modulator) or network to digital device (demodulator).
- **Half duplex** − A half-duplex modem has the capacity to transfer data in both the directions but only one at a time.
- **Full duplex** − A full duplex modem can transmit data in both the directions simultaneously.

Uniform Resource Locator

A **uniform resource locator (URL)** is a reference to a resource that specifies the location of the resource on a computer network and a mechanism for retrieving it.

A URL is a specific type of uniform resource identifier (URI), although many people use the two terms interchangeably. A URL implies the means to access an indicated resource, which is not true of every URI. URLs occur most commonly to reference web pages (http), but are also used for file transfer (ftp), email (mailto), database access (JDBC), and many other applications.

Most web browsers display the URL of a web page above the page in an address bar. A typical URL has the form *http://www.example.com/index.html*, which indicates the protocol type (*http*), the domain name, (*www.example.com*), and the specific web page (*index.html*).

History

The Uniform Resource Locator was standardized in 1994 by Tim Berners-Lee and the URI working group of the Internet Engineering Task Force (IETF) as an outcome of collaboration started at the IETF Living Documents “Birds of a Feather” session in 1992. The format

combines the pre-existing system of domain names (created in 1985) with file path syntax, where slashes are used to separate directory and file names. Conventions already existed where server names could be prepended to complete file paths, preceded by a double-slash (//).

Berners-Lee later regretted the use of dots to separate the parts of the domain name within URIs, wishing he had used slashes throughout. For example,*http://www.example.com/path/to/name* would have been written *http:com/example/www/path/to/name*. Berners-Lee has also said that, given the colon following the URI scheme, the two slashes before the domain name were also unnecessary.

Syntax

Every HTTP URL consists of the following, in the given order. Several schemes other than HTTP also share this general format, with some variation.

- the scheme name (commonly called protocol, although not every URL scheme is a protocol, e.g. mailto is not a protocol)
- a colon, two slashes,
- a host, normally given as a domain name For example, *http://www.example.com/path/to/name* would have been written *http:com/example/www/path/to/name* but sometimes as a literal IP address
- optionally a colon followed by a port number
- the full path of the resource

The scheme says *how* to connect, the host specifies *where* to connect, and the remainder specifies *what* to ask for.

For programs such as Common Gateway Interface (CGI) scripts, this is followed by a query string, and an optional fragment identifier.

The syntax is:

scheme://[user:password@]domain:port/path?query_string#fragment_id

Component details:

- The **scheme**, which in many cases is the name of a protocol (but not always), defines how the resource will be obtained. Examples include http, https, ftp, file and many others. Although schemes are case-insensitive, the canonical form is lowercase.

- The **domain name** or literal numeric IP address gives the destination location for the URL. A literal numeric IPv6 address may be given, but must be enclosed in *[ ]* e.g.*[db8:0cec::99:123a]*.

  The domain *google.com*, or its numeric IP address *173.194.34.5*, is the address of Google's website.

- The domain name portion of a URL is not case sensitive since DNS ignores case:

  http://en.example.org/ and *HTTP://EN.EXAMPLE.ORG/* both open the same page.

- The **port number**, given in decimal, is optional; if omitted, the default for the scheme is used.

  For example, *http://vnc.example.com:5800* connects to port 5800 of vnc.example.com, which may be appropriate for a VNC remote control session. If the port number is omitted for an http: URL, the browser will connect on port 80, the default HTTP port. The default port for an https: request is 443.

- The **path** is used to specify and perhaps find the resource requested. This **path** may or may not describe folders on the file system in the web server. It may be very different from the arrangement of folders on the web server. It is case-sensitive, though it may be treated as case-insensitive by some servers, especially those based on Microsoft Windows.

  If the server is case sensitive and *http://en.example.org/wiki/URL* is correct,then *http://en.example.org/WIKI/URL* or *http://en.example.org/wiki/url* will display an HTTP 404 error page, unless these URLs point to valid resources themselves.

- The **query string** contains data to be passed to software running on the server. It may contain name/value pairs separated by ampersands, for example

  ?first_name=John&last_name=Doe.

- The **fragment identifier**, if present, specifies a part or a position within the overall resource or document.

  When used with HTML, it usually specifies a section or location within the page, and used in combination with Anchor elements or the "id" attribute of an element, the browser is scrolled to display that part of the page.

The scheme name defines the namespace, purpose, and the syntax of the remaining part of the URL. Software will try to process a URL according to its scheme and context. For example, a web browser will usually dereference the URL *http://example.org:80* by performing an HTTP request to the host at *example.org*, using port number 80.

Other examples of scheme names include https, gopher, wais, ftp. URLs with https as a scheme (such as *https://example.com/*) require that requests and responses will be made over a secure connection to the website. Some schemes that require authentication allow a username, and perhaps a password too, to be embedded in the URL, for example*ftp://asmith@ftp.example.org*. Passwords embedded in this way are not conducive to security, but the full possible syntax is,

scheme://username:password@domain:port/path?query_string#fragment_id

Other schemes do not follow the HTTP pattern. For example, the *mailto* scheme only uses valid email addresses. When clicked on in an application, the URL *mailto:bob@example.com*may start an e-mail composer with the address *bob@example.com* in the To field. The *tel* scheme is even more different; it uses the public switched telephone network for addressing, instead of domain names representing Internet hosts.

List of allowed URL characters

Unreserved

The alphanumerical upper and lower case character may optionally be encoded:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9 – _ . ~

Reserved

Special symbols must sometimes be percent-encoded:

! * ‘ ( ) ; : @ & = + $, / ? % # [ ]

Further details can for example be found in RFC 3986 and http://www.w3.org/Addressing/URL/uri-spec.html.

Relationship to URI

See also: URIs, Relationship to URL and URN

A URL is a URI that, in addition to identifying a web resource, provides a means of locating the resource by describing its “primary access mechanism (e.g., its network location)”.

IP ADDRESS

An IP address is a unique address that identifies a device on the internet or a local network. IP stands for "Internet Protocol," which is the set of rules governing the format of data sent via the internet or local network.

In essence, IP addresses are the identifier that allows information to be sent between devices on a network: they contain location information and make devices accessible for communication. The internet needs a way to differentiate between different computers, routers, and websites. IP addresses provide a way of doing so and form an essential part of how the internet works.

What is an IP?

An IP address is a string of numbers separated by periods. IP addresses are expressed as a set of four numbers — an example address might be 192.158.1.38. Each number in the set can range from 0 to 255. So, the full IP addressing range goes from 0.0.0.0 to 255.255.255.255.

IP addresses are not random. They are mathematically produced and allocated by the Internet Assigned Numbers Authority (IANA), a division of the Internet Corporation for Assigned Names and Numbers (ICANN). ICANN is a non-profit organization that was established in the United States in 1998 to help maintain the security of the internet and allow it to be usable by all. Each time anyone registers a domain on the internet, they go through a domain name registrar, who pays a small fee to ICANN to register the domain.

How do IP addresses work

If you want to understand why a particular device is not connecting in the way you would expect or you want to troubleshoot why your network may not be working, it helps understand how IP addresses work.

Internet Protocol works the same way as any other language, by communicating using set guidelines to pass information. All devices find, send, and exchange information with other connected devices using this protocol. By speaking the same language, any computer in any location can talk to one another.

The use of IP addresses typically happens behind the scenes. The process works like this:

1. Your device indirectly connects to the internet by connecting at first to a network connected to the internet, which then grants your device access to the internet.
2. When you are ~~at~~ home, that network will probably be your Internet Service Provider (ISP). At work, it will be your company network.
3. Your IP address is assigned to your device by your ISP.
4. Your internet activity goes through the ISP, and they route it back to you, using your IP address. Since they are giving you access to the internet, it is their role to assign an IP address to your device.
5. However, your IP address can change. For example, turning your modem or router on or off can change it. Or you can contact your ISP, and they can change it for you.
6. When you are out and about – for example, traveling – and you take your device with you, your home IP address does not come with you. This is because you will be using another network (Wi-Fi at a hotel, airport, or coffee shop, etc.) to access the internet and will be using

a different (and temporary) IP address, assigned to you by the ISP of the hotel, airport or coffee shop.

As the process implies, there are different types of IP addresses, which we explore below.

Types of IP addresses

There are different categories of IP addresses, and within each category, different types.

Consumer IP addresses

Every individual or business with an internet service plan ~~will~~ have two types of IP addresses: their private IP addresses and their public IP address. The terms public and private relate to the network location — that is, a private IP address is used inside a network, while a public one is used outside a network.

Private IP addresses

Every device that connects to your internet network has a private IP address. This includes computers, smartphones, and tablets but also any Bluetooth-enabled devices like speakers, printers, or smart TVs. With the growing internet of things, the number of private IP addresses you have at home is probably growing. Your router needs a way to identify these items separately, and many items need a way to recognize each other. Therefore, your router generates private IP addresses that are unique identifiers for each device that differentiate them on the network.

Public IP addresses

A public IP address is the primary address associated with your whole network. While each connected device has its own IP address, they are also included within the main IP address for your network. As described above, your public IP address is provided to your router by your ISP. Typically, ISPs have a large pool of IP addresses that they distribute to their customers. Your public IP address is the address that all the devices outside your internet network will use to recognize your network.

Public IP addresses

Public IP addresses come in two forms – dynamic and static.

Dynamic IP addresses

Dynamic IP addresses change automatically and regularly. ISPs buy a large pool of IP addresses and assign them automatically to their customers. Periodically, they re-assign them and put the older IP addresses back into the pool to be used for other customers. The rationale for this approach is to generate cost savings for the ISP. Automating the regular movement of IP addresses means they don't have to carry out specific actions to re-establish a customer's IP address if they move home, for example. There are security benefits, too, because a changing IP address makes it harder for criminals to hack into your network interface.

Static IP addresses

In contrast to dynamic IP addresses, static addresses remain consistent. Once the network assigns an IP address, it remains the same. Most individuals and businesses do not need a static IP address, but for businesses that plan to host their own server, it is crucial to have one. This is because a static IP address ensures that websites and email addresses tied to it will have a consistent IP address — vital if you want other devices to be able to find them consistently on the web.

This leads to the next point – which is the two types of website IP addresses.

There are two types of website IP addresses

For website owners who don't host their own server, and instead rely on a web hosting package – which is the case for most websites – there are two types of website IP addresses. These are shared and dedicated.

Shared IP addresses

Websites that rely on shared hosting plans from web hosting providers will typically be one of many websites hosted on the same server. This tends to be the case for individual websites or SME websites, where traffic volumes are manageable, and the sites themselves are limited in terms of the number of pages, etc. Websites hosted in this way will have shared IP addresses.

Dedicated IP addresses

Some web hosting plans have the option to purchase a dedicated IP address (or addresses). This can make obtaining an SSL certificate easier and allows you to run your own File Transfer Protocol (FTP) server. This makes it easier to share and transfer files with multiple people within an organization and allow anonymous FTP sharing options. A dedicated IP address also allows you to access your website using the IP address alone rather than the domain name — useful if you want to build and test it before registering your domain.

IP address security threats

Cybercriminals can use various techniques to obtain your IP address. Two of the most common are social engineering and online stalking.

Attackers can use social engineering to deceive you into revealing your IP address. For example, they can find you through Skype or a similar instant messaging application, which uses IP addresses to communicate. If you chat with strangers using these apps, it is important to note that they can see your IP address. Attackers can use a Skype Resolver tool, where they can find your IP address from your username.

How to protect and hide your IP address

Hiding your IP address is a way to protect your personal information and online identity. The two primary ways to hide your IP address are:

1. Using a proxy server
2. Using a virtual private network (VPN)

A proxy server is an intermediary server through which your traffic is routed:

- The internet servers you visit see only the IP address of that proxy server and not your IP address.
- When those servers send information back to you, it goes to the proxy server, which then routes it to you.

A drawback of proxy servers is that some of the services can spy on you — so you need to trust it. Depending on which one you use, they can also insert ads into your browser.

VPN offers a better solution:

- When you connect your computer – or smartphone or tablet – to a VPN, the device acts as if it is on the same local network as the VPN.
- All your network traffic is sent over a secure connection to the VPN.
- Because your computer behaves as if it is on the network, you can securely access local network resources even when you are in another country.
- You can also use the internet as if you were present at the VPN's location, which has benefits if you are using public Wi-Fi or want to access geo-blocked websites.

Kaspersky Secure Connection is a VPN that protects you on public Wi-Fi, keeps your communications private, and ensures that you are not exposed to phishing, malware, viruses, and other cyber threats.

Other ways to protect your privacy

Change privacy settings on instant messaging applications

Apps installed on your device are a major source of IP address hacking. Instant messaging and other calling apps can be used as a tool by cybercriminals. Using IM apps only allows direct connections from contacts and doesn't accept calls or messages from people you don't know. Changing your privacy settings makes it harder to find your IP address because people who don't know you cannot connect with you.

Create unique passwords

Your device password is the only barrier that can restrict people from accessing your device. Some people prefer to stick to their devices' default passwords, which makes them vulnerable to attack. Like all your accounts, your device needs to have a unique and strong password that is not easy to decode. A strong password contains a mix of upper- and lower-case letters, numerals, and characters. This will help to safeguard your device against IP address hacking.

Stay alert to phishing emails and malicious content

A high proportion of malware and device tracking software is installed via phishing emails. When you connect with any site, this provides the site with access to your IP address and device location, making it vulnerable to hacking. Be vigilant when opening emails from

unknown senders and avoid clicking on links that could send you to unauthorized sites. Pay close attention to the emails' content, even if they appear to come from well-known sites and legitimate businesses.

Use a good antivirus solution and keep it up to date

Install comprehensive antivirus software and keep it up to date. For example, Kaspersky's Anti-Virus protection guards you from viruses on your PC and Android devices, secures and stores your passwords and private documents, and encrypts the data you send and receive online with VPN.

Protecting your IP address is a crucial aspect of protecting your online identity. Securing it through these steps is a way to stay safe against the wide variety of cybercriminals' attacks.

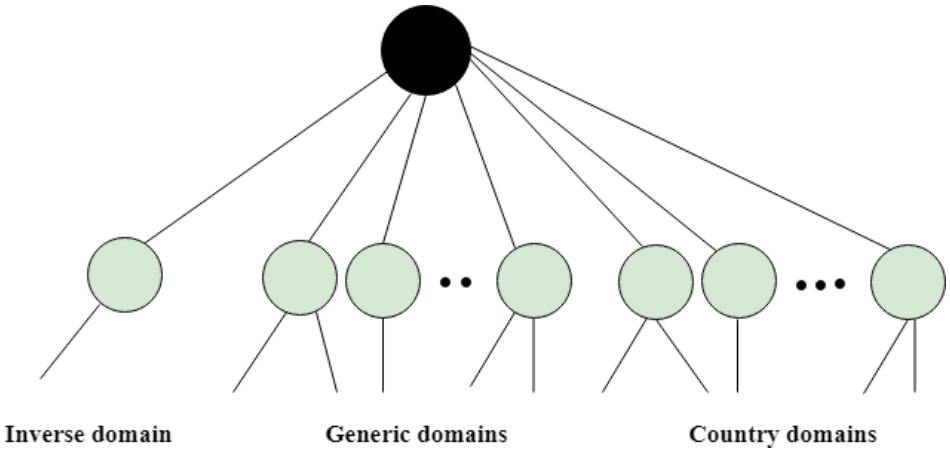
DOMAIN NAME SYSTEM

The Domain Name System (DNS) is the Internet's system for mapping alphabetic names to numeric Internet Protocol (IP) addresses like a phone book maps a person's name to a phone number. For example, when a Web address (URL) is typed into a browser, a DNS query is made to learn an IP address of a Web server associated with that name.

Using the *www.example.com* URL, *example.com* is the domain name, and *www* is the hostname. DNS resolution maps www.example.com into an IP address (such as 192.0.2.1). When a user needs to load a webpage, a conversion must occur between what a user types into their web browser (*www.example.com*) into an IP address required to locate the *www.example.com* site.

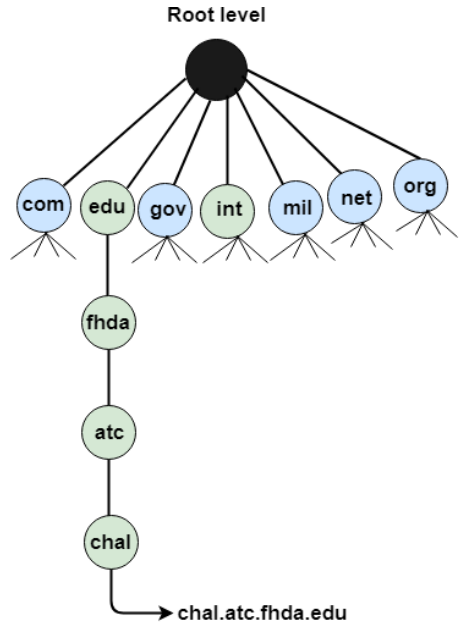
The DNS system is an open worldwide network of database name servers that include 13 authoritative name servers that serve the DNS root zone level, known as "root servers". A root server (also called a DNS root nameserver) receives a DNS query that includes a domain name (e.g. *www.thousandeyes.com*), and responds by directing that request to a top-level domain (TLD) nameserver, based on the TLD of that domain such as .com, .net, and .org. It directly responds to requests for DNS records in the root zone by returning an appropriate list of the authoritative TLD name servers for the appropriate TLD that can resolve the initial DNS lookup request for an IP address of that domain name.

DNS is a TCP/IP protocol used on different platforms. The domain name space is divided into three different sections: generic domains, country domains, and inverse domain.



Generic Domains

- It defines the registered hosts according to their generic behavior.
- Each node in a tree defines the domain name, which is an index to the DNS database.
- It uses three-character labels, and these labels describe the organization type.

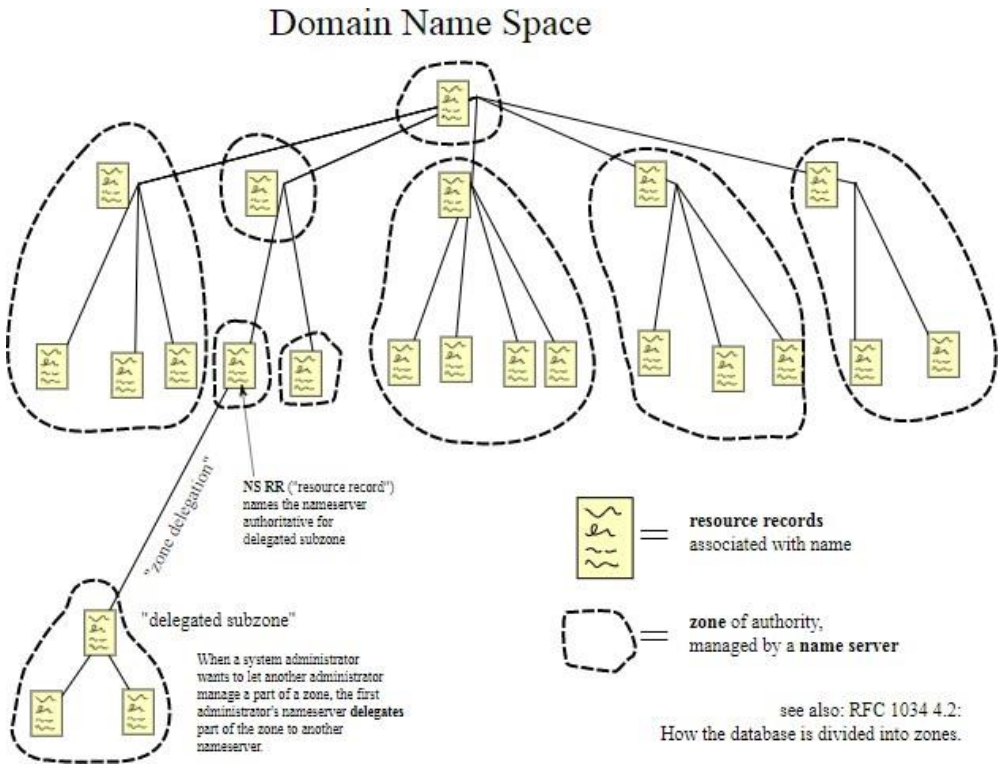


Domain name space

The domain name space consists of a tree data structure. Each node or leaf in the tree has a *label* and zero or more *resource records* (RR), which hold information associated with the domain name. The domain name itself consists of the label, concatenated with the name of its parent node on the right, separated by a dot.

The tree sub-divides into *zones* beginning at the root zone. A DNS zone may consist of only one domain, or may consist of many domains and sub-domains, depending on the administrative choices of the zone manager. DNS can also be partitioned according to *class* where the separate classes can be thought of as an array of parallel namespace trees.

Administrative responsibility for any zone may be divided by creating additional zones. Authority over the new zone is said to be *delegated* to a designated name server. The parent zone ceases to be authoritative for the new zone.



What is an Authoritative DNS Server

Authoritative DNS servers are the DNS infrastructure that satisfies requests from recursive DNS servers (discussed below) with the corresponding IP address information. Authoritative DNS servers also provide essential DNS information for each website (corresponding IP addresses, a list of mail servers and other DNS record information).

An authoritative DNS server holds and maintains DNS records. It is the last server in a DNS lookup chain that responds with the queried DNS record. An authoritative DNS ultimately allows a web browser with the URL request to reach the IP address needed to access a website or other web resources. An authoritative DNS domain name server is a definitive source for DNS domain name resolution.

DNS security technology is used to protect DNS information stored as a record in the Domain Name System (DNS). It provides secure authentication for the origin of the DNS data, helping to safeguard against attacks and protect data integrity.

What is a Recursive DNS Server

All domains are assigned a unique IP address on the Internet. When a website address is typed into a browser, like *google.com*, the browser response is to convert this URL into the correct IP address for this website. The web browser starts this process by utilizing an internal cache of recent DNS query results. This cache is the first place the browser checks (if it has this capability) to find the IP address of the requested domain. If this does not result in a DNS resolution, a client-side DNS resolver sends a DNS query to a recursive DNS server that could reside at an Internet Service Provider (ISP) or public DNS provider.

Every DNS record has a TTL or time-to-live parameter that specifies how long a recursive DNS server can cache it. To enhance site performance it is important to reduce DNS lookups. A DNS server cannot be used efficiently without a properly configured TTL. For this reason a DNS TTL check can be extremely important in assessing loading speeds. If the DNS recursive server has the DNS record cached or stored for some time as the TTL parameter specifies, then it answers the DNS query by providing the cached source or IP information (recursive lookup).

If the DNS record is not in the recursive DNS server's cache, it queries the root DNS server for the Top Level Domain of the site the user/client is trying to reach (in this example,

google.com). The Root DNS server then responds with a pointer to forwards the DNS lookup request to the TLD nameserver that identifies the authoritative DNS server that is responsible for returning the corresponding site IP address of the website enabling the browser to access the desired website.

Working of DNS

- DNS is a client/server network communication protocol. DNS clients send requests to the. server while DNS servers send responses to the client.
- Client requests contain a name which is converted into an IP address known as a forward DNS lookups while requests containing an IP address which is converted into a name known as reverse DNS lookups.
- DNS implements a distributed database to store the name of all the hosts available on the internet.
- If a client like a web browser sends a request containing a hostname, then a piece of software such as **DNS resolver** sends a request to the DNS server to obtain the IP address of a hostname. If DNS server does not contain the IP address associated with a hostname, then it forwards the request to another DNS server. If IP address has arrived at the resolver, which in turn completes the request over the internet protocol.

WEB SERVERS

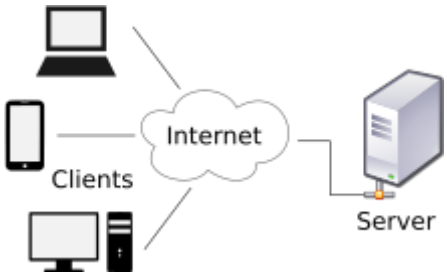
A **web server** is computer software and underlying hardware that accepts requests via HTTP (the network protocol created to distribute web content) or its secure variant HTTPS. A user agent, commonly a web browser or web crawler, initiates communication by making a request for a web page or other resource using HTTP, and the server responds with the content of that resource or an error message. A web server can also accept and store resources sent from the user agent if configured to do so.

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.

The hardware used to run a web server can vary according to the volume of requests that it needs to handle. At the low end of the range are embedded systems, such as a router that runs a small web server as its configuration interface. A high-traffic Internet website might handle requests with hundreds of servers that run on racks of high-speed computers.

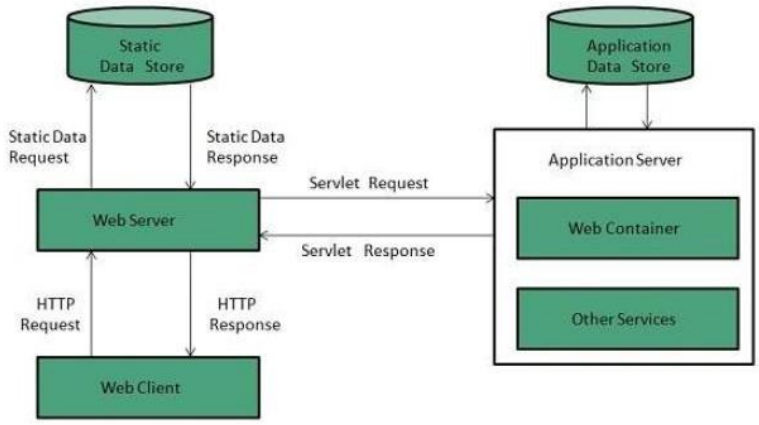
A resource sent from a web server can be a preexisting file (static content) available to the web server, or it can be generated at the time of the request (dynamic content) by another program that communicates with the server software. The former usually can be served faster and can be more easily cached for repeated requests, while the latter supports a broader range of applications.

Technologies such as REST and SOAP, which use HTTP as a basis for general computer-to-computer communication, as well as support for WebDAV extensions, have extended the application of web servers well beyond their original purpose of serving human-readable pages.



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.



Key Points

- 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.

Architecture

Web Server Architecture follows the following two approaches:

1. Concurrent Approach
2. Single-Process-Event-Driven Approach.

Concurrent Approach

Concurrent approach allows the web server to handle multiple client requests at the same time. It can be achieved by following methods:

- Multi-process
- Multi-threaded
- Hybrid method.

Multi-processing

In this a single process (parent process) initiates several single-threaded child processes and distribute incoming requests to these child processes. Each of the child processes are responsible for handling single request.

It is the responsibility of parent process to monitor the load and decide if processes should be killed or forked.

Multi-threaded

Unlike Multi-process, it creates multiple single-threaded process.

Hybrid

It is combination of above two approaches. In this approach multiple process are created and each process initiates multiple threads. Each of the threads handles one connection. Using multiple threads in single process results in less load on system resources.

APACHE

Apache HTTP Server is a free and open-source web server that delivers web content through the internet. It is commonly referred to as Apache and after development, it quickly became the most popular HTTP client on the web. It's widely thought that Apache gets its name from its development history and process of improvement through applied patches and modules but that was corrected back in 2000. It was revealed that the name originated from the respect of the Native American tribe for its resiliency and durability.

Now, before we get too in depth on Apache, we should first go over what a web application is and the standard architecture usually found in web apps.

Apache Web Application Architecture

Apache is just one component that is needed in a web application stack to deliver web content. One of the most common web application stacks involves LAMP, or Linux, Apache, MySQL, and PHP.

Linux is the operating system that handles the operations of the application. Apache is the web server that processes requests and serves web assets and content via HTTP. MySQL is the database that stores all your information in an easily queried format. PHP is the programming language that works with apache to help create dynamic web content.

While actual statistics may vary, it's fair to say a large portion of web applications run on some form of the LAMP stack because it is easy to build and also free to use. For the most part, web applications tend to generally have similar architecture and structure even though they serve many different functions and purposes. Most web applications also benefit from Firewalls, Load Balancers, Web Servers, Content Delivery Networks, and Database Servers.

Firewalls help protect the web application from both external threats and internal vulnerabilities depending on where the firewalls are configured. Load Balancers help distribute traffic across the web servers which handle the HTTP(S) requests (this is where

Apache comes in) and application servers (servers that handle the functionality and workload of the web app.) We also have Database Servers, which handle asset storage and backups. Depending on your infrastructure, your database and application can both live on the same server although it's recommended to keep those separate.

Apache is considered open source software, which means the original source code is freely available for viewing and collaboration. Being open source has made Apache very popular with developers who have built and configured their own modules to apply specific functionality and improve on its core features. Apache has been around since 1995 and is responsible as a core technology that helped spur the initial growth of the internet in its infancy.

One of the pros of Apache is its ability to handle large amounts of traffic with minimal configuration. It scales with ease and with its modular functionality at its core, you can configure Apache to do what you want, how you want it. You can also remove unwanted modules to make Apache more lightweight and efficient.

Some of the most popular modules that can be added are SSL, Server Side Programming Support (PHP), and Load Balancing configs to handle large amounts of traffic. Apache can also be deployed on Linux, MacOS, and Windows. If you learn how to configure Apache on Linux, you can administer Apache on Windows and Mac. The only difference would be directory paths and installation processes.

Features of Apache Web Server

- Handling of static files
- Loadable dynamic modules
- Auto-indexing
- .htaccess
- Compatible with IPv6
- Supports HTTP/2
- FTP connections
- Gzip compression and decompression
- Bandwidth throttling
- Perl, PHP, Lua scripts

- Load balancing
- Session tracking
- URL rewriting
- Geolocation based on IP address

Apache Web Server

Throughout the last few decades, Apache has proven to be a staple in many popular stacks and the backbone of the early internet year. While it's popularity is declining and the options of web server choices are increasing, Apache still plays a pivotal role in many technology stacks and companies system infrastructure. Even with new technologies and servers coming out nonstop, Apache is still a technology every developer should learn how to handle and configure.

Small Questions

S. No	Questions	LOCF Mapping
1.	What is the World Wide Web? Who invented it?	K1
2.	Define Invisible Web.	K1
3.	Differentiate between a router and a modem.	K2
4.	What is TCP/IP?	K1
5.	Expand URL, DNS, and ISP.	K1

Big Questions

S. No	Questions	LOCF Mapping
1.	Discuss the evolution and growth of the World Wide Web from its inception to the present.	K2
2.	Explain the Internet architecture with its hardware and software components.	K2
3.	Describe the client-server principle. How does it work in web communication?	K2, K3
4.	Explain the different types of internet connections available today.	K2
5.	What is a URL? Explain its syntax and components with examples.	K2, K3

UNIT II: Web Design Principles

Content: Web Design Principles; Markup Languages (SGML, HTML, XML); Web Browsers (Internet Explorer, Mozilla Firefox, Google Chrome); Communication Tools (Email, Instant Messaging, VoIP, Video Conferencing).

UNIT – 2

WEB DESIGN PRINCIPLES:

website design often comes from different expert areas. It's not a one-step process, you can never design good websites from shortcut steps. After taking a graduate class named "***Human-Computer Communication***" we had a long discussion in the class after which we came up with the 9 general principles with the help of which you can improve or design good websites (It's like a formula of design). We should always focus on these key aspects i.e. ***Usability*** and ***User Experience*** for designing a website. Follow the below discussed 9 general principles and make your websites usable, aesthetic, friendly, engaging to use.

1. Easy Navigation

This is one of the most important features or a principle that you must keep in mind and implement while you are designing a website. According to the survey conducted by ***Clutch,*** they found that "***Almost everyone (94%) says easy navigation is the most important website feature***". It's true because if your site is hard to navigate, nothings easily available in the first place, then things might not go smoothly. ***Easy navigation includes a simple menu layout and the ability to quickly and reliably move through sites***. Don't worry I have some tips for easy website navigation.

- Keep the ***navigation bar*** as simple as possible with ***minimal options***
- ***Mitigate*** the drop-down menu options
- Follow ***real-world conventions*** or use the user's language for naming the options. Don't use jargons
- Avoid ***too many clicks*** inside the website

Motivate, ***learn*** and ***teach*** yourself how to implement the above tips by looking at some of the best websites available out there. For example, **Gmail** — Google.

2. Responsive design

First, let's get the basics ready. Responsive design means the design of a website/webpage that fits well with all the smartphones, computers, laptops or any display devices irrespective of their aspect ratios. An aspect ratio is the ***height*** and ***width*** of a display device. Every user today wants a mobile version of the website. It is the duty of the designer not only to design the website for a bigger screen but also focus on the smaller screens too. The website should not work only for a specific iPhone, Blackberry (if that's a thing today), Samsung and One Plus phones. The design should fit every single smartphone available in the market. Well, if not every smartphone, at least try to. Some tips for responsive design are:

- ***Optimize*** the images
- Ensure ***buttons*** can be easily clicked on smaller screens
- Create ***several prototypes***
- Consider a ***Mobile-first design*** approach

Hint*: If you need your to design to be the best then follow:*

Understand →Explore →Prototype →Evaluate

Now you guys might have heard this term before a million times "***Mobile-first design approach***". There are some design requirements in this approach which are listed below:

- Providing ***less content*** on the screen and focus on important details
- The ***text size*** should be appropriate
- Provide ***better interactive elements***
- Telling the ***customers*** you want them
- ***Client Squish Media*** (keeping the aspect ratio in mind while designing)

3. Same Color Scheme (Consistency)

When defining knowledge hierarchy through the websites, color is very important. Users should be able to skim through pages understand what they are about. Keeping the coloring scheme consistent is one of the hardest tasks on the planet. Sometimes the color which we like might not be liked by others. There is often a tradeoff between colors. Make sure that the color you choose is well-liked by others. Some colors might fit well to a certain text and some might not, so maintaining consistency is the key role here. Run a survey, test, and iterate until you get good feedback from the users. Also, the color scheme on your website must be consistent. Don't do a combo of colors. That's a bad idea. Below is the chart of the psychology of choosing colors.

Some of the tips are:

- Don't use ***super bright*** or ***dark colors*** on your website.
- ***Highlight*** the important information where necessary.
- Make sure you use the ***right color combination***.
- Keep the colors ***minimalistic*** as possible.

4. Comfortable UI

A user interface is through which the user interacts with the system. It serves as a bridge between the system and the user. If the UI (User Interface) is good then the user would definitely like to spend more time on it. It is the job of the designer to make the UI look fresh and clear. Below are some tips to design a good UI:

- Keep the ***interface*** simple
- Utilize the ***page layout*** efficiently
- Consistent ***fonts*** and ***color***
- Remove ***irrelevant*** information
- Avoid ***infinite scrolling*** of information

To make something intuitive, connect it to users' own experiences. Use Metaphors while designing. Often the three most important questions that we need to ask ourselves before we design something are:

- ***Who*** the users are
- ***What*** activities are being carried out
- ***Where*** the interaction is taking place.

We need to optimize the interactions users have with their product. So they match the user's activities and needs.

5. Contents meet user goals for visiting websites

When it comes to finding the right information, a user would not be happy if he/she doesn't get what they want. Be specific in providing the information. For example, the ***Yale University School of Arts*** website, under the tuition fees section they have just provided "***The tuition fee for the academic year 2019–2020 is $39,924. The Corporation of Yale University reserves the right to revise tuition rates as necessary".***

Consider an international student and his mom/dad is exploring this website because they are curious about the tuition fees. They accidentally stumbled upon this page and they are clueless about this vague information. Now the questions that can arise in their mind are:

- ***Why*** is the tuition fees $39, 924
- ***What*** is the international student fees
- ***Where*** is the proper tuition fees breakdown

See, this is how the user feels frustrated, in this case, the contents never meet user goals. Another scenario is the ***help and documentation*** pages are such pages that need more attention while providing the details. Even if the program can be used without paperwork, support and documents may be required. Any such information should be:

- ***Easy*** to search
- Be ***specific***
- ***Focus*** on the user's task
- List as ***concrete steps*** to be carried out

Hint: *Whatever information you lay to users remember "****Never beat around the bush****"*

6. Performance: Quick to show something

The performance of the website must be ***smooth as butter,*** not ***slow as a sloth*** (that's a terrible example). If the performance of the website is too slow, then it's will create a bad impact on the company of the website. And the uses might stop using the website. Think and act wisely. There are many website testing tools out there, just choose one and test the speed of your website. One of the popular ones is ***Pingdom Tools.*** So the first step is rather than updating your website without knowing where the problem is. Try one of those tools and they get to where the problem is, which page, section or whatever's performance is slow. The correct that

specific problem. Things might make more sense later on. Some tips to increase the performance of your website is given down below:

- ***Compress*** your files
- ***Optimize*** your images
- Avoid using many images: ***use text*** instead
- ***Reduce*** HTTP requests

I tested the Gmail website using the Pingdom Tools the results were fascinating. Obviously, it's google's product, the end results would be amazing.

7. Feedback about progress

Feedback is a response about the task, process or event after its completion or during its stage of completion. Know what's going on inside the system or website in this case. Consider you are using a website for doing some transactions online, there's one point of time where you don't know what going on in the system. Suddenly you get a message of transaction failed. At this point you are clueless. It would have been better if the website provides appropriate feedback about the transaction failure or the error behind it such as internet connection issues, money limit in your bank, or even a slight website's technical glitch. In fact, the feedback about progress is an important aspect of design it's also the first ***usability heuristics of user-interface design***.

Suppose when I upload an image as an attachment on Gmail, with the help of the progress bar I can come to know that the image is being uploaded. These simple things make your website more functional. Some tips for providing feedback about the progress are as given below:

- Use appropriate ***loading*** or ***percentages*** to show about the progress
- If an item or a product is ***not available*** for purchase, let the user know about it.

- Provide a ***success*** or a ***failure*** screen, such that users come to know that they are on the right track.

8. Avoid “Alert/Dialogs” when not necessary

Don’t annoy the user’s by providing unnecessary dialogs every time when they do something on the website. For example, think the user is trying to make a payment online. Provide a dialog or alert only when the transaction is successful or failure. Don’t provide alerts/dialogue every time when they enter their card details such as name, DOB, or the pin number. Also, sometimes it would be good if you ask the user to receive the notifications before displaying them without their permission.

Some of the things that you need to follow are:

- If at all you provide the dialogs or alert, be ***specific*** don’t fill it with irrelevant information.
- ***Remove*** all the dialogs/alerts where unnecessary
- Don’t just annoy the user by the ***opening the dialogs*** on the website
- Don’t put the user inside the ***loop*** of dialogs

9. Try to mitigate 404 or 500 errors.

If you have already designed a website or on the verge of designing, then try to develop it completely with all the pages, features, and functionality: the websites must not have broken links and images, incomplete functionalities and many more. Especially the site should not have any **404 or 500 errors**. If at all you have broken links by mistake, then try to design a good 404 error page. A user must be happy even though they see a 404 error page but this case is exceptional. Obviously, when your website is almost perfect you don’t have to worry about it. For example, I have been using google’s website for quite a long I have never faced such an issue. One of the best pages for 404 errors in terms of design is the ***GitHub website***. It’s fully animated too. Check it out.

The above are the ***9 General Principles for Good Website Design.*** With the help of these principles, you can easily develop user-friendly and functional websites. Without these fundamentals, it would certainly be hard to develop a good and intuitive website. Just keep in mind a website with good user-friendly and usability will always succeed in the real world.

WEB DESIGN PRINCIPLES MARK-UP LANGUAGES:

The method of publishing hypertexts on the web employs a **markup language** called **Hypertext Markup Language** or **HTML** for short. It's perhaps no wonder that HTML was developed in Switzerland, a country with four official languages. It may have been an awareness of the difficulties involved with understanding different languages that led developers at the CERN research facility to invent a sort of Esperanto for computers, because HTML is platform independent and understood by every computer system.

Apart from HTML, there are a number of other markup languages for many different purposes (e.g. Markdown, DocBook, SVG, LaTeX or PostScript). The main feature of markup languages, and HTML in particular, is the usage of plain text files in which the structure and formatting information is directly integrated into the text by means of **tags**. Originally, markup such as this was used in the text as instructions for typesetters in the printing industry. Nowadays, an **interpreter** (hardware or software) handles the conversion of these instructions into visual information.

The table below illustrates the principle of tags using as examples HTML and another markup language designed for typesetting created by *Leslie Lamport* called **LaTeX** (pronounced *Lah-tekh*).

Example	HTML	LaTeX	Possible rendering of the output
Heading	`<h1>Text</h1>`	`\section{Text}`	**Text**

Bullet lists	`<ul>`
	`<li>Item 1</li>`
	`<li>Item 2</li>`
	`<li>Item 3</li>`
	`</ul>`
		`\begin{itemize}`
		`\item Item 1`
		`\item Item 2`
		`\item Item 3`
		`\end{itemize}`
			• Item 1
			• Item 2
			• Item 3

Bold text	`<b>Text</b>`	`\bf{Text}`	**Text**
Italic text	`<i>Text</i>`	`\i{Text}`	*Text*

Markup languages are currently being used to create documents in many applications: In the printing industry for instance, the **PostScript** markup language developed by Adobe is used to print page layouts at variable sizes on many different output devices. Things like the font, font size, indentation, margins, size and position of images, etc. are defined using PostScript. PostScript enabled printers and presses are equipped with a special interpreter that converts PostScript instructions into visual information. Over the years, PostScript has become a standard in the printing industry, but is being increasingly pushed aside by PDF, also developed by Adobe.

SGML:

SGML (Standard Generalized Markup Language)

SGML (Standard Generalized Markup Language) is a standard for how to specify a document markup language or tag set. Such a specification is itself a document type definition (DTD). SGML is not in itself a document language, but a description of how to specify one. It is metadata.

SGML is based on the idea that documents have structural and other semantic elements that can be described without reference to how such elements should be displayed. The actual display of such a document may vary, depending on the output medium and style preferences. Some advantages of documents based on SGML are:

- They can be created by thinking in terms of document structure rather than appearance characteristics (which may change over time).
- They will be more portable because an SGML compiler can interpret any document by reference to its document type definition (DTD).
- Documents originally intended for the print medium can easily be re-adapted for other media, such as the computer display screen.

The language that this Web browser uses, Hypertext Markup Language (HTML), is an example of an SGML-based language. There is a document type definition for HTML (and reading the HTML specification is effectively reading an expanded version of the document type definition). In today's distributed networking environment, many documents are being described with the Extensible Markup Language (XML) which is a data description language (and a document can be viewed as a collection of data) that uses SGML principles.

SGML is used for marking up documents and has the advantage of not being dependent on a specific application. It is derived from GML (generalized markup language), which allowed users to work on standardized formatting styles for electronic documents.

SGML is based somewhat on earlier generalized markup languages developed at IBM, including General Markup Language (GML) and ISIL.

Standard generalized markup language features the following characteristics:

- Descriptive Markup
- Document Types

Descriptive markup involves the use of markup code that identify how various portions of a document should be interpreted. For example, the code may identify one portion as a paragraph, another as a footnote and still another as a list or an item in a list.

Any software capable of processing the marked-up document will then do so using its own kind of rendering. For example, one application might gather portions identified as footnotes and print them out at the end of each page. Another might print footnotes at the end of each chapter. Still another might not print out the footnotes at all.

Another important characteristic of standard generalized markup language is its use of document types, and subsequently it use of document type definition (DTD). A particular document type is expected to have specific parts and a specific structure. For example, when there is a DTD for a report, the portions and structure of the document should follow what is defined in the DTD for it to be considered a report. One major benefit is that documents with the same type can be processed uniformly by all software capable of processing them.

HTML

HTML is the **language in which most websites are written**. HTML is used to create pages and make them functional. HTML was first created by Tim Berners-Lee, Robert Cailliau, and others starting in **1989**. It stands for Hyper Text Markup Language. Hypertext means that the document contains **links that allow the reader to jump to other places** in the document or to another document altogether. The latest version is known as HTML5.

A **Markup Language** is a way that computers speak to each other to control how text is processed and presented. To do this HTML uses two things: tags and **attributes**.

Tags and attributes are the basis of HTML.

They work together but perform different functions – it is worth investing 2 minutes in **differentiating the two**.

What Are HTML Tags?

Tags are used to **mark up the start of an HTML element** and they are usually enclosed in angle brackets. An example of a tag is: `<h1>`.

Most tags must be opened `<h1>` and closed `</h1>` in order to function.

What are HTML Attributes?

Attributes contain **additional pieces of information**. Attributes take the form of an opening tag and additional info is **placed inside**.

An example of an attribute is:

`<img src="mydog.jpg" alt="A photo of my dog.">`

In this instance, the image source (src) and the alt text (alt) are attributes of the `<img>` tag.

How To Add Text In HTML

Adding text to our HTML page is simple using an element opened with the tag <p> which **creates a new paragraph**. We place all of our regular text inside the element <p>.

When we write text in HTML, we also have a number of other elements we can use to **control the text or make it appear in a certain way**. Other Key Elements They are as follows:

Element	Meaning	Purpose
<b>	Bold	Highlight important information
****	Strong	Similarly to bold, to highlight key text
<i>	Italic	To denote text
****	Emphasised Text	Usually used as image captions
<mark>	Marked Text	Highlight the background of the text
<small>	Small Text	To shrink the text
<strike>	Striked Out Text	To place a horizontal line across the text
<u>	Underlined Text	Used for links or text highlights
<ins>	Inserted Text	Displayed with an underline to show an inserted text

Element	Meaning	Purpose
<sub>	Subscript Text	Typographical stylistic choice
<sup>	Superscript Text	Another typographical presentation style

These tags **must** be opened and closed around the text in question.

Let's try it out. On a new line in the HTML editor, type the following HTML code:

Welcome to my brand new website. This site will be my new home on the web.

Don't forget to **hit save and then refresh the page** in your browser to see the results.

HTML Data types

HTML defines several data types for element content, such as script data and stylesheet data, and a plethora of types for attribute values, including IDs, names, URIs, numbers, units of length, languages, media descriptors, colors, character encodings, dates and times, and so on. All of these data types are specializations of character data.

SGML-based versus XML-based HTML

One difference in the latest HTML specifications lies in the distinction between the SGML-based specification and the XML-based specification. The XML-based specification is usually called XHTML to distinguish it clearly from the more traditional definition. However, the root element name continues to be "html" even in the XHTML-specified HTML. The W3C intended XHTML 1.0 to be identical to HTML 4.01 except where limitations of XML over the more complex SGML require workarounds. Because XHTML and HTML are closely related, they are sometimes documented in parallel. In such circumstances, some authors conflate the two names as (X)HTML or X(HTML).

Like HTML 4.01, XHTML 1.0 has three sub-specifications: strict, transitional and frameset.

Aside from the different opening declarations for a document, the differences between an HTML 4.01 and XHTML 1.0 document—in each of the corresponding DTDs—are largely syntactic. The underlying syntax of HTML allows many shortcuts that XHTML does not, such as elements with optional opening or closing tags, and even empty elements which must not have an end tag. By contrast, XHTML requires all elements to have an opening tag and a closing tag. XHTML, however, also introduces a new shortcut: an XHTML tag may be

opened and closed within the same tag, by including a slash before the end of the tag like this: `<br/>`. The introduction of this shorthand, which is not used in the SGML declaration for HTML 4.01, may confuse earlier software unfamiliar with this new convention. A fix for this is to include a space before closing the tag, as such: `<br />`.[94]

To understand the subtle differences between HTML and XHTML, consider the transformation of a valid and well-formed XHTML 1.0 document that adheres to Appendix C (see below) into a valid HTML 4.01 document. To make this translation requires the following steps:

1. **The language for an element should be specified with `lang` attribute rather than the XHTML `xml:lang` attribute.** XHTML uses XML's built in language-defining functionality attribute.
2. **Remove the XML namespace (`xmlns=URI`).** HTML has no facilities for namespaces.
3. **Change the document type declaration** from XHTML 1.0 to HTML 4.01. (see DTD section for further explanation).
4. If present, **remove the XML declaration.** (Typically this is: `<?xml version="1.0" encoding="utf-8"?>`).
5. **Ensure that the document's MIME type is set to `text/html`.** For both HTML and XHTML, this comes from the HTTP `Content-Type` header sent by the server.
6. **Change the XML empty-element syntax to an HTML style empty element** (`<br />` to `<br>`).

Those are the main changes necessary to translate a document from XHTML 1.0 to HTML 4.01. To translate from HTML to XHTML would also require the addition of any omitted opening or closing tags. Whether coding in HTML or XHTML it may just be best to always include the optional tags within an HTML document rather than remembering which tags can be omitted.

A well-formed XHTML document adheres to all the syntax requirements of XML. A valid document adheres to the content specification for XHTML, which describes the document structure.

The W3C recommends several conventions to ensure an easy migration between HTML and XHTML (see HTML Compatibility Guidelines). The following steps can be applied to XHTML 1.0 documents only:

- Include both `xml:lang` and `lang` attributes on any elements assigning language.
- Use the empty-element syntax only for elements specified as empty in HTML.
- Include an extra space in empty-element tags: for example `<br />` instead of `<br>`.
- Include explicit close tags for elements that permit content but are left empty (for example, `<div></div>`, not `<div />`).
- Omit the XML declaration.

By carefully following the W3C's compatibility guidelines, a user agent should be able to interpret the document equally as HTML or XHTML. For documents that are XHTML 1.0 and have been made compatible in this way, the W3C permits them to be served either as HTML (with a `text/html` MIME type), or as XHTML (with an `application/xhtml+xml` or `application/xml` MIME type). When delivered as XHTML, browsers should use an XML parser, which adheres strictly to the XML specifications for parsing the document's contents.

XML:

XML stands for **Extensible Mark-up Language**, developed by W3C in 1996. It is a text-based mark-up language derived from Standard Generalized Mark-up Language (SGML). XML 1.0 was officially adopted as a W3C recommendation in 1998. XML was designed to carry data, not to display data. XML is designed to be self-descriptive. XML is a subset of SGML that can define your own tags. A Meta Language and tags describe the content. XML Supports CSS, XSL, DOM. XML does not qualify to be a programming language as it does not performs any computation or algorithms. It is usually stored in a simple text file and is processed by special software that is capable of interpretingXML.

The Difference between XML and HTML

1. HTML is about displaying information, where asXML is about carrying information. In other words, XML was created to structure, store, and transport information. HTML was designed to display thedata.

2. Using XML, we can create own tags where as in HTML it is not possible instead it offers several built intags.

3. XML is platform independent neutral and languageindependent.

4. XML tags and attribute names are case-sensitive where as in HTML it isnot.

5. XML attribute values must be single or double quoted where as in HTML it is not compulsory.

6. XML elements must be properlynested.

7. All XML elements must have a closingtag.

Well Formed XML Documents

A "Well Formed" XML document must have the following correct XML syntax:

- XML documents must have a rootelement
- XML elements must have a closing tag(start tag must have matching endtag).
- XML tags are casesensitive
- XML elements must be properly nestedEx:<one><two>Hello</two></one>
- XML attribute values must bequoted

XML with correct syntax is "Well Formed" XML. XML validated against a DTD is "Valid" XML.

What is Markup?

XML is a markup language that defines set of rules for encoding documents in a format that is both human-readable andmachine-readable.

Example for XML Document

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?><!—xml declaration-->
<note>
<to>MRCET</to>
<from>MRGI</from>
 <heading>KALPANA</heading>
 <body>Hello, world! </body>
</note>
```

□□Xml document begins with XML declaration statement: <? xml version="1.0" encoding="ISO-8859-1"?>.

 The next line describes the **root element** of the document:**<note>**.

 This element is "the parent" of all otherelements.

 The next 4 lines describe 4**child elements** of the root: to, from, heading, and body. And finally the last line defines the end of the root element : **</note>.**

 The XML declaration has no closing tag i.e.</?xml>

 The **default standalone value** is set to **no**. Setting it to **yes** tells the processor there are no external declarations (DTD) required for parsing the document. The file name extension used for xml program is.xml.

Valid XML document

If an XML document is well-formed and has an associated Document Type Declaration (DTD), then it is said to be a valid XML document. We will study more about DTD in the chapter XML - DTDs.

XML DTD

Document Type Definition purpose is to define the structure of an XML document. It defines the structure with a list of defined elements in the xml document. Using DTD we can specify the various elements types, attributes and their relationship with one another. Basically DTD is used to specify the set of rules for structuring data in any XML file.

Why use a DTD?

XML provides an application independent way of sharing data. With a DTD, independent groups of people can agree to use a common DTD for interchanging data. Your application can use a standard DTD to verify that data that you receive from the outside world is valid. You can also use a DTD to verify your own data.

DTD - XML building blocks

Various building blocks of XML are-

1. Elements: The basic entity is **element**. The elements are used for defining the tags. The elements typically consist of opening and closing tag. Mostly only one element is used to define a singletag.

2. Tags

Tags are used to markup elements. A starting tag like <element_name> mark up the beginning of an element, and an ending tag like </element_name> mark up the end of an element.

Examples:

A body element: <body>body text in between</body>. A message element: <message>some message in between</message>

3. Attribute: The attributes are generally used to specify the values of the element. These are specified within the double quotes. Ex: <flagtype=‖true‖>

4. Entities

Entities as variables used to define common text. Entity references are references to entities. Most of you will known the HTML entity reference: " " that is used to insert an extra space in an HTML document. Entities are expanded when a document is parsed by an XML parser.

The following entities are predefined in XML:

< (<), >(>), &(&), "(") and '(').

5. CDATA: It stands for character data. CDATA is text that will **NOT be parsed by a parser**. Tags inside the text will NOT be treated as markup and entities will not beexpanded.

6. PCDATA: It stands for Parsed Character Data(i.e., text). Any parsed character data should not contain the markup characters. The markup characters are < or > or &. If we want to use these characters then make use of < , > or &. Think of character data as the text found between the start tag and the end tag of an XML element. PCDATA is text that will be **parsed by a parser**

XML Schemas

 XML Schema is commonly known as XML Schema Definition (XSD). It is used to describe and validate the structure and the content of XML data. XML schema defines the elements, attributes and data types. Schema element supports Namespaces. It is similar to a database schema that describes the data in a database. XSD extension is**".xsd"**.

 This can be used as an alternative to XML DTD. The XML schema became the W#C recommendation in2001.

 XML schema defines elements, attributes, element having child elements, order of child elements. It also defines fixed and default values of elements andattributes.

 XML schema also allows the developer to us **datatypes**.

XML Parsers

An XML parser converts an XML document into an XML DOM object - which can then be manipulated with a JavaScript.

Two types of XML parsers:

ValidatingParser

• It requires document type declaration

• It generates error if document doesnot

- o Conform with DTDand
- o Meet XML validityconstraints

Non-validating Parser

• It checks well-formedness for xmldocument

• It can ignore externalDTD

What is XML Parser? XML Parser provides way how to access or modify data present in an XML document. Java provides multiple options to parse XML document. Following are various types of parsers which are commonly used to parse XML documents.

Types of parsers:

Dom Parser - Parses the document by loading the complete contents of the document and creating its complete hiearchical tree inmemory.

SAX Parser - Parses the document on event based triggers. Does not load the complete document into thememory.

JDOM Parser - Parses the document in similar fashion to DOM parser but in more easier way.

StAX Parser - Parses the document in similar fashion to SAX parser but in more efficient way.

XPath Parser - Parses the XML based on expression and is used extensively in conjuction withXSLT.

DOM4J Parser - A java library to parse XML, XPath and XSLT using Java Collections Framework , provides support for DOM, SAX andJAXP.

DOM-Document Object Model

The Document Object Model protocol converts an XML document into a collection of objects in your program. XML documents have a hierarchy of informational units called nodes; this hierarchy allows a developer to navigate through the tree looking for specific information. Because it is based on a hierarchy of information, the DOM is said to be tree based. DOM is a way of describing those nodes and the relationships between them.

You can then manipulate the object model in any way that makes sense. This mechanism is also known as the "random access" protocol, because you can visit any part of the data at any time. You can then modify the data, remove it, or insert new data.

The XML DOM, on the other hand, also provides an API that allows a developer to add, edit, move, or remove nodes in the tree at any point in order to create an application. A DOM parser creates a tree structure in memory from the input document and then waits for requests from client. A DOM parser always serves the client application with the **entire document no matter how much is actually needed** by the client. With DOM parser, method calls in client application have to be explicit and forms a kind of chained method calls.

Document Object Model is for defining the standard for accessing and manipulating XML documents. **XML DOM** is used for

- **Loading the xmldocument**
- **Accessing the xmldocument**
- **Deleting the elements of xmldocument**
- **Changing the elements of xml document**

WEB Servers:

A **web browser** (commonly referred to as a **browser**) is a software application for retrieving, presenting and traversing information resources on the World Wide Web. An *information resource* is identified by a Uniform Resource Identifier (URI/URL) and may be a web page, image, video or other piece of content. Hyperlinks present in resources enable users easily to navigate their browsers to related resources.

Although browsers are primarily intended to use the World Wide Web, they can also be used to access information provided by web servers in private networks or files in file systems.

The major web browsers are Firefox, Internet Explorer, Google Chrome, Opera, and Safari.

The first web browser was invented in 1990 by Sir Tim Berners-Lee. Berners-Lee is the director of the World Wide Web Consortium (W3C), which oversees the Web's continued development, and is also the founder of the World Wide Web Foundation. His browser was called WorldWideWeb and later renamed Nexus.

The first commonly available web browser with a graphical user interface was Erwise. The development of Erwise was initiated by Robert Cailliau.

In 1993, browser software was further innovated by Marc Andreessen with the release of Mosaic, "the world's first popular browser", which made the World Wide Web system easy to use and more accessible to the average person. Andreesen's browser sparked the internet boom of the 1990s. The introduction of Mosaic in 1993 – one of the first graphical web browsers – led to an explosion in web use. Andreessen, the leader of the Mosaic team at National Center for Supercomputing Applications (NCSA), soon started his own company, named Netscape, and released the Mosaic-influenced Netscape Navigator in 1994, which quickly became the world's most popular browser, accounting for 90% of all web use at its peak (see usage share of web browsers).

Microsoft responded with its Internet Explorer in 1995, also heavily influenced by Mosaic, initiating the industry's first browser war. Bundled with Windows, Internet Explorer gained dominance in the web browser market; Internet Explorer usage share peaked at over 95% by 2002.

Opera debuted in 1996; it has never achieved widespread use, having less than 2% browser usage share as of February 2012 according to Net Applications. Its Opera-mini version has an additive share, in April 2011 amounting to 1.1% of overall browser use, but focused on the fast-growing mobile phone web browser market, being preinstalled on over 40 million phones. It is also available on several other embedded systems, including Nintendo's Wii video game console.

In 1998, Netscape launched what was to become the Mozilla Foundation in an attempt to produce a competitive browser using the open source software model. That browser would eventually evolve into Firefox, which developed a respectable following while still in the beta stage of development; shortly after the release of Firefox 1.0 in late 2004, Firefox (all versions) accounted for 7% of browser use. As of August 2011, Firefox has a 28% usage share.

Apple's Safari had its first beta release in January 2003; as of April 2011, it had a dominant share of Apple-based web browsing, accounting for just over 7% of the entire browser market.

The most recent major entrant to the browser market is Chrome, first released in September 2008. Chrome's take-up has increased significantly year by year, by doubling its usage share from 8% to 16% by August 2011.

This increase seems largely to be at the expense of Internet Explorer, whose share has tended to decrease from month to month. In December 2011, Chrome overtook Internet Explorer 8 as the most widely used web browser but still had lower usage than all versions of Internet Explorer combined. Chrome's user-base continued to grow and in May 2012, Chrome's usage passed the usage of all versions of Internet Explorer combined. By April 2014, Chrome's usage had hit 45%.

Internet Explorer will be deprecated in Windows 10, with Microsoft Edge replacing it as the default web browser.

To view and browse pages on the Web, really what you want is an internet browser. To distribute pages on the Web, you want a web server. A web server is the program that runs on a computer and is liable for answering to internet browser demands for records. You want a web server to distribute archives on the Web. Whenever you utilize a program to demand a page on a site, that program makes a web association with a server utilizing the HTTP convention. The browser then that the formats point to arranges the data it got from the server. Server acknowledges the association, sends the substance of the mentioned records and afterward closes.

WEB Browsers:

WWW Clients, or " Browsers ": The program you can use and to get the WWW is known as a program since it " browses " the WWW and solicitations these hypertext reports. Programs can be graphical, permits to see and hear the illustrations and sound; text-just programs (i.e., those with no sound or designs ability) are additionally accessible. These projects get http and other Internet conventions like FTP, gopher, mail and news, making the WWW a sort of "one quit shopping" for Internet users.

An web browser is the program you can use to see pages and explore the World Wide Web. A wide exhibit of internet browsers is accessible for pretty much in every stage you can envision. Microsoft Internet Explorer, for instance, is incorporated with Windows and Safari is included with Mac OS X. Mozilla Firefox, Netscape Navigator, and Opera are all accessible free of charge. A **web browser** is a type of software that allows you to find and view websites on the Internet. Even if you didn't know it, you're using a web browser right now to read this page! There are many different web browsers, but some of the most common ones include **Google Chrome**, **Internet Explorer**, **Safari**, **Microsoft Edge** and **Mozilla Firefox**.

No matter which web browser you use, you'll want to learn the basics of browsing the Web. N**avigating** to different websites, **using tabbed browsing**, creating **bookmarks**, and more.

History of Web Browser

Today web browsers are easily accessible and can be used on devices like computer, laptops, mobile phones, etc. but this evolution of making browsers available for easy use took many years.

Given below are some salient points which one must know with regard to the history of web browsers:

- **“WorldWideWeb”** was the first web browser created by Tim Berners Lee in 1990. This is completely different from the World Wide Web we use today
- In 1993, the **“Mosaic”** web browser was released. It had the feature of adding images and an innovative graphical interface. It was the “the world’s first popular browser”
- After this, in 1994, Marc Andreessen (leader of Mosaic Team) started working on a new web browser, which was released and was named **“Netscape Navigator”**

- In 1995, **"Internet Explorer"** was launched by Microsoft. It soon overtook as the most popular web browser
- In 2002, **"Mozilla Firefox"** was introduced which was equally as competent as Internet Explorer
- Apple too launched a web browser in the year 2003 and named it **"Safari"**. This browser is commonly used in Apple devices only and not popular with other devices
- Finally, in the year 2008, Google released "Chrome" and within a time span of 3 years it took over all the other existing browsers and is one of the most commonly used web browsers across the world

Functions of Web Browser

Our dependency on the Internet has massively increased. Stated below are functions of web browsers and how are they useful:

- The main function is to retrieve information from the World Wide Web and making it available for users
- Visiting any website can be done using a web browser. When a URL is entered in a browser, the web server takes us to that website
- To run Java applets and flash content, plugins are available on the web browser
- It makes Internet surfing easy as once we reach a website we can easily check the hyperlinks and get more and more useful data online
- Browsers user internal cache which gets stored and the user can open the same webpage time and again without losing extra data
- Multiple webpages can be opened at the same time on a web browser
- Options like back, forward, reload, stop reload, home, etc. are available on these web browsers, which make using them easy and convenient

Types of Web Browser

The functions of all web browsers are the same. Thus, more than the different types there are different web browsers which have been used over the years.

Discussed below are different web browser examples and their specific features:

1. World Wide Web

- The first web browser ever
- Launched in 1990
- It was later named “Nexus” to avoid any confusion with the World Wide Web
- Had the very basic features and less interactive in terms of graphical interface
- Did not have the feature of bookmark

2. Mosaic

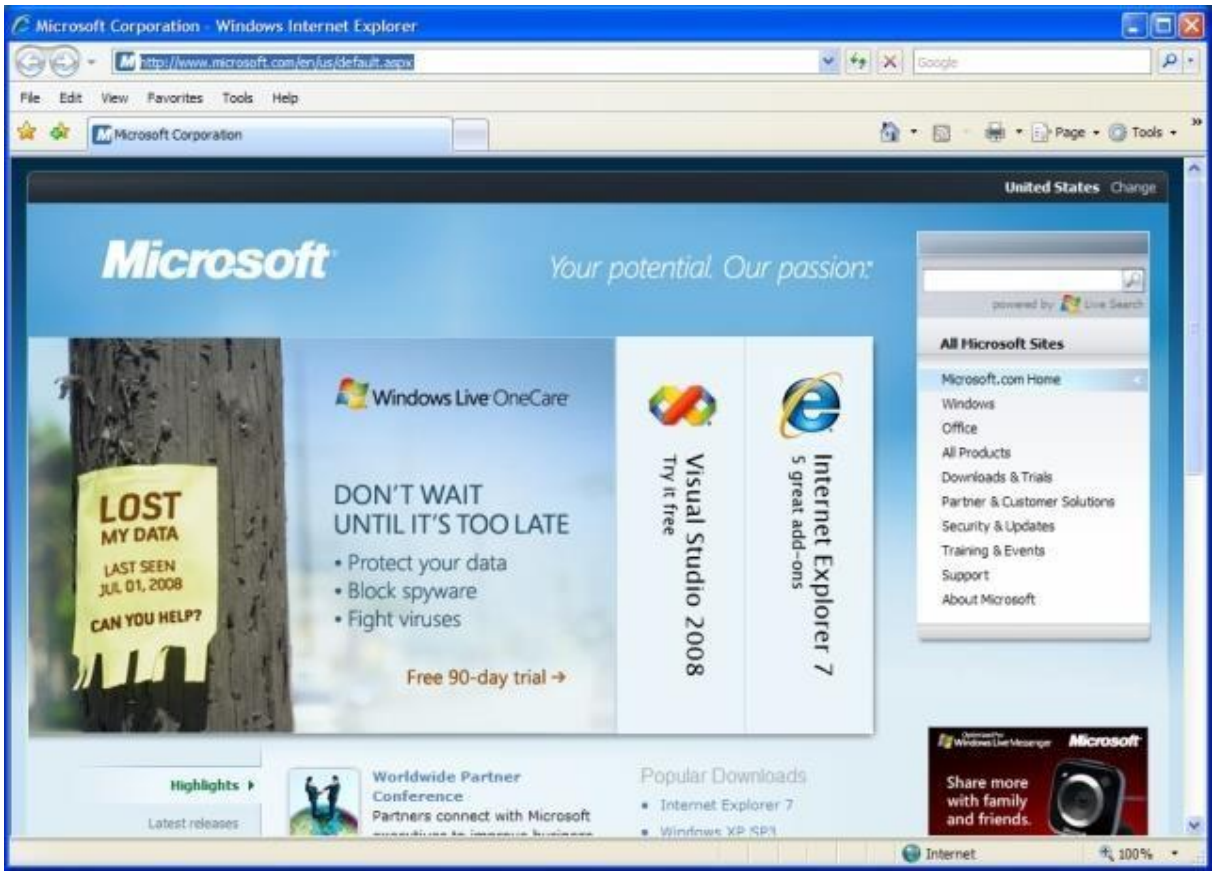
- It was launched in 1993
- The second web browser which was launched
- Had a better graphical interface. Images, text and graphics could all be integrated
- It was developed at the National Center for Supercomputing Applications
- The team which was responsible for creating Mosaic was lead by Marc Andreessen
- It was named “the world’s first popular browser”

3. Netscape Navigator

- It was released in 1994
- In the 1990s, it was the dominant browser in terms of usage share
- More versions of this browser were launched by Netscape
- It had an advanced licensing scheme and allowed free usage for non-commercial purposes

4. Internet Explorer

Internet Explorer (IE - created by Microsoft) is a very prominant web browser for the Windows OS. IE is the most popular web browser. It comes pre-installed on all Windows computers. The latest version of IE is IE7 with IE8 in beta. IE was designed to view a broad range of web pages and to provide certain features within the OS.



IE almost fully supports HTML 4.01, CSS Level 1, XML 1.0, and DOM Level 1. It has introduced a number of proprietary extensions to many of the standards. This has resulted in a number of web pages that can only be viewed properly using IE. It has been subject to many security vulnerabilities just like Windows has. Much of the spyware, adware, and viruses across the Internet are made possible by exploitable bugs and flaws in the security architecture of IE. These are were drive-by downloads come into play (see computer security lesson for more details on that)

- It was launched in 1995 by Microsoft
- By 2003, it has attained almost 95% of usage share and had become the most popular browsers of all
- Close to 10 versions of Internet Explorer were released by Microsoft and were updated gradually
- It was included in the Microsoft Windows operating system
- In 2015, it was replaced with “Microsoft Edge”, as it became the default browser on Windows 10

5. Firefox

Firefox is a very popular web browser. One of the great things about Firefox is that it is supported on all different OSs. Firefox is also open source which makes its support group a very large community of open source developers. Firefox is also known for its vast range of plugins/add-ons that let the user customize in a variety of ways. Firefox is a product of the Mozilla Foundation. The latest version of Firefox is Firefox 3.

Some of Firefox's most prominant features include: tabbed browsing, a spell checker, incremental find, live bookmarking, a download manager, and an integrated search system that uses the user's favorite search engine. Like mentioned before, one of the best things about Firefox is its vast amount of plugins/add-ons. Some of the most popular include NoScript (script blocker), FoxyTunes (controls music players), Adblock Plus (ad blocker), StumbleUpon (website discovery), DownThemAll! (download functions), and Web Developer (web tools).

- It was introduced in 2002 and was developed by Mozilla Foundation
- Firefox overtook the usage share from Internet Explorer and became the dominant browser during 2003-04
- Location-aware browsing was made available with Firefox
- This browser was also made available for mobile phones, tablets, etc.

6. Google Chrome

- It was launched in 2008 by Google

- It is a cross-platform web browser
- Multiple features from old browsers were amalgamated to form better and newer features
- To save computers from malware, Google developed the ad-blocking feature to keep the user data safe and secure
- Incognito mode is provided where private searching is available where no cookies or history is saved
- Till date, it has the best user interface

Apart from these, Opera Mini web browser was introduced in 2005 which was specially designed for mobile users. Before the mobile version, the computer version "Opera" was also released in 1995. It supported a decent user interface and was developed by Opera Software.

COMMUNICATION TOOLS:

In previous years, you learned that digital or electronic communication refers to any data, information, words or photos that are sent electronically in order to communicate with one or more people. This includes calls, messages, group chats, emails, social networks and websites. These methods of electronic communication have made our lives much easier and enabled us to communicate with people around the world. However, they do have some flaws and risks that you need to be aware of.

ADVANTAGES OF ELECTRONIC COMMUNICATION

- **Speedy transmission**: It requires only a few seconds to communicate through electronic media because it supports quick transmission.
- **Wide coverage**: The world has become a global village and communication around the globe requires only a second.
- **Low cost**: Electronic communication saves time and money. For example, text SMS is cheaper than the traditional letter.
- **Exchange of feedback**: Electronic communication allows the instant exchange of feedback. So communication becomes perfect using electronic media.
- **Managing global operation**: Due to the advancement of electronic media, business managers can easily control operation across the globe. Video or

teleconferencing e-mail and mobile communication are helping managers in this regard.

DISADVANTAGES OF ELECTRONIC COMMUNICATION

- The volume of data: The volume of telecommunication information is increasing at such a fast rate that business people are unable to absorb it within the relevant time limit.
- The cost of development: Electronic communication requires huge investment for infrastructural development. Frequent change in technology also demands further investment.
- Legal status: Data or information, if faxed, may be distorted and will cause zero value in the eye of law.
- Undelivered data: Data may not be retrieved due to system error or fault with the technology. Hence required service will be delayed.
- Dependency: Technology is changing every day and therefore poor countries face the problem as they cannot afford the new or advanced technology. Therefore poor countries need to be dependent towards developed countries for sharing global network.

When you communicate on the internet – whether it is via email, instant messaging, or posting a blog – it is important that you follow proper **netiquette**. This will not only make the internet a more pleasant place for everyone else, it will also save you from potential embarrassment in the future!

GUIDELINES WHEN COMMUNICATING ON THE INTERNET

- Texting (messaging):
  - Keep texts short
  - Longer texts can be misinterpreted
  - Sign a text with your name
  - Spell out all words and do not use “texting lingo” or shorthand
  - Texts can be saved and can be altered
- Email:
  - Use a descriptive subject line
  - Be courteous

- Reply promptly – but allow yourself time to get over an initial reaction to an angry email
- Remember attachments to an email may contain metadata that can disclose unwanted information to the recipient

- Social Media:
  - There is no expectation of privacy on the internet
  - Change your passwords frequently
  - Log off after visiting the page
  - Delete your browsing history, saved passwords and cookies regularly
  - Do not disparage anyone via social media
  - Educate yourself about a site before joining.

While the tips covered in this section are generally good guidelines, it is important to note that netiquette differs from site to site and changes over time. Find out what is acceptable behaviour on that website before sending your own messages.

TYPES OF ELECTRONIC COMMUNICATION

With the advances in computer technology and the internet, there are many new and exciting ways to communicate; from sending instant messages on social network sites, to email. The most popular types of digital communications, their advantages, as well as their disadvantages are as follows,

EMAIL

Email is one of the first and most popular forms of electronic communication. It allows the user to send and receive files and messages over the internet, and can be used on a wide variety of devices. Here are some of the advantages and disadvantages of email.

ADVANTAGES OF EMAIL

- Email is a free tool.
- Email is quick. Once you have finished composing a message, sending it is as simple as clicking a button. Once it is sent and delivered, it can be read almost immediately.
- Email is simple. It is easy to use, email allows for the easy and quick access of information and contacts.
- Email allows for easy referencing. Messages that have been sent and received can stored, and searched through safely and easily.
- Email is accessible from anywhere – as long as you have an internet connection.
- Email is paperless, and therefore, beneficial for the planet.
- Email allows for mass sending of messages, you can send one particular message to several recipients all at once.
- Email allows for instant access of information and files.

DISADVANTAGES OF EMAIL

- Email could potentially cause information overload. Some messages may be dismissed or left unread, especially if there are a lot coming in and the network has not integrated some sort of email alert system into the computers at work.
- Email lacks a personal touch.
- Email can be disruptive. Going through each email can be disruptive to work as it does require a bit of time. This disruption is decreased through the utilisation of an email alert system.
- Email cannot be ignored for a long time.

- Emails can cause misunderstandings. Because email does not include nonverbal communication, recipients may misinterpret the sender's message. This is particularly true of senders fail to go through their messages before they send them.
- Email messages can contain viruses. It's best to be aware of this possibility so that you are careful when opening messages from people you don't know, or when downloading attachments.
- Email should be kept short and brief. This is especially difficult if you are one to send messages that are too long.
- Email requires timely responses. While some people tend to disregard messages, those that require responses should be replied to as soon as they are received and read. If not, urgent and important messages may be left untended.

When communicating by email, always note the following:

- Be polite.
- Do not write emails that are unnecessarily long.
- Do not spam people with emails that they are not interested in.
- Make sure that your message is logical and says what you intended to say.

Email is not limited to only sending messages over the internet; it provides users with many features. We will now take a look at each of these in some more detail.

INSTANT MESSAGING

Instant messaging refers to short messages that are sent in real time over the internet. The messages can include multimedia items, such as pictures, videos and voice recordings.

ADVANTAGES OF INSTANT MESSAGING

- Messages are free to send
- Messages are received directly after being sent
- You can see if the message has been delivered
- You can see when your message has been read
- You can send a variety of messages; including text messages, pictures, videos, music and web links
- You can create group conversations in order to discuss a specific topic or plan events

DISADVANTAGES OF INSTANT MESSAGING

- Messages are not always saved
- It is an informal method of communication and might not be suited for business-related communications
- There is a pressure to respond immediately as people can see when you read their messages
- Can be distracting as one message can lead to a whole conversation
- Low security, as instant messaging services use a public network

When communicating by instant messaging, take note of the following:

- Do not expect an immediate reply. The person you are messaging might be busy and will reply once he or she is available.
- Keep your messages short and to the point.
- Do not type your messages using uppercase as it can be interpreted as shouting.
- Be polite.
- Do not use slang words and abbreviations. This might save you time, but it can also confuse others if they are not aware of the meaning.

VOIP

VoIP is a type of digital communication that allows the user to speak with one or more users over the internet. This type of communication is very similar to a phone call, with the

exception being that it uses your internet connection and, therefore, uses data. Here are some of the advantages and disadvantages of VoIP.

ADVANTAGES OF VOIP

- VoIP is much cheaper than using traditional telecommunication services
- It can help with productivity as you can have face-to-face meetings with colleagues in different cities
- It saves time and money as you can have face-to-face business meetings without having to travel to the client

DISADVANTAGES OF VOIP

- You need an active internet connection
- Audio quality depends on the quality of your internet connection
- Some VoIP programs use large amounts of data

When communicating using VoIP, take note of the following:

- Be polite and speak clearly.
- Indicate that you are listening to the person with whom you are speaking.
- Repeat important details to ensure that you understand them correctly.
- Do not be afraid to ask questions if something is not clear.

VIDEO CONFERENCING

- Everyone can see you, all the time. This is not an audio conference, just because you are not speaking does not mean others in the conference can't still see you.
- Be punctual and courteous. Introduce yourself and take note of other attendees' names so you can address them by name. Turn off ringers for your other phones. Treat this just like you would an on-site meeting.
- No multi-tasking, we can see you. Look at your screen, pay attention to others and when speaking make sure to look at your camera.
- If it is improper for a face-to-face meeting, then it doesn't work for video either. Don't click your pen, tap on your desk or anything else annoying or distracting. Avoid yawning, gum chewing, etc.
- Make sure you have good light. Adjust lighting or use a portable light source to make sure you have good light shining at you from the front. You can overdo it too, so

experiment until you find a good balance. Try pointing a strong desk lamp at the wall you're facing. You get good front light without having to look directly into a harsh light.

- Do not eat! You may enjoy a glass of water or coffee, but drinking a glass of wine or a bottle of beer is not acceptable. Do not smoke during the meetings either.
- Video allows us to do face-to-face meetings right from our virtual office. Even though we all enjoy sitting in a short and t-shirt in our virtual offices, it is not appropriate when you are called into a virtual meeting. Business casual at all times is the new rule. If you have a planned customer call you should consider dressing the same as you would for an on-site meeting.
- Do video calls from your desk or other appropriate location. Lying on the couch (or anywhere) with your pc on your chest or stomach doesn't present a flattering view.
- If you have a cluttered work space, make sure it's not showing up on camera. Consider removing award plaques from other (competitor) companies on your wall.
- Make sure to have current client version loaded before scheduled calls. Test your audio and/or video before a scheduled call.
- Avoid high traffic areas. Sometimes it's hard to avoid but do not position yourself where other people will be parading through the view of the camera on a regular basis.
- Close unused applications, video can be CPU/memory intensive.
- Avoid creating pixilation! Do not wear stripes, or anything with a heavy pattern. If you have vertical mini-blinds do not have them in the background. Minimise your hand gestures and body/head movements as well.
- Consider posting your comment/question in the chat window.
- Picture in Picture is your best reference, you can see yourself and your surroundings just as others on the call can. Pay close attention to what you see there, and make adjustments as necessary.
- DO NOT video while driving.

Small Questions

S. No	Questions	LOCF Mapping
1.	List any five principles of good website design.	K1
2.	What is the difference between HTML and XML?	K2
3.	Expand SGML. What is its significance?	K1
4.	Name four popular web browsers.	K1
5.	What is VoIP?	K1

Big Questions

S. No	Questions	LOCF Mapping
1.	Discuss the nine general principles of good website design.	K2, K3
2.	Explain the features and applications of HTML and XML markup languages.	K2
3.	Describe the history, functions, and types of web browsers.	K2
4.	Discuss the advantages and disadvantages of email, instant messaging, and VoIP as communication tools.	K2, K3
5.	Compare and contrast HTML, SGML, and XML.	K4

UNIT III: Social Media and Web Development

Content: Social Media – Definition, Business Applications, Benefits, Challenges, Types, Examples; Web Development Technologies (Browsers, HTML & CSS, Frameworks, Programming Languages, Protocols, API, Data Formats); Design in the Browser; Web Hosting and Publishing

UNIT-3

SOCIAL MEDIA:

Social media is a collective term for websites and applications that focus on communication, community-based input, interaction, content-sharing and collaboration.

People use social media to stay in touch and interact with friends, family and various communities. Businesses use social applications to market and promote their products and track customer concerns.

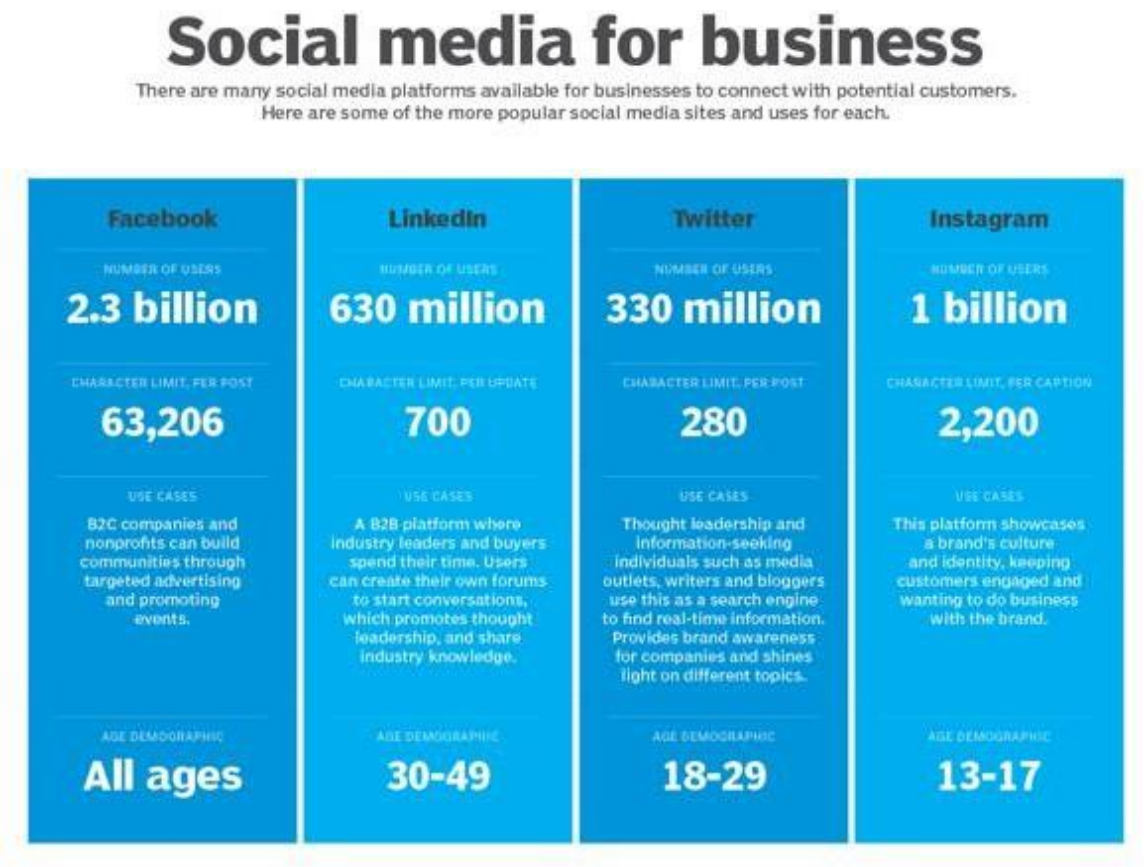
Business-to-consumer websites include social components, such as comment fields for users. Various tools help businesses track, measure and analyze the attention the company gets from social media, including brand perception and customer insight.

Social media has enormous traction globally. Mobile applications make these platforms easily accessible. Some popular examples of general social media platforms include Twitter, Facebook and LinkedIn.

What are the business applications of social media?

In business, social media is used to market products, promote brands, connect to customers and foster new business. As a communication platform, social media promotes customer feedback and makes it easy for customers to share their experiences with a company. Businesses can respond quickly to positive and negative feedback, address customer problems and maintain or rebuild customer confidence.

Social media is also used for crowdsourcing. That's the practice of using social networking to gather knowledge, goods or services. Companies use crowdsourcing to get ideas from employees, customers and the general public for improving products or developing future products or services.



Benefits of social media

Social media provides several benefits, including the following:

- **User visibility.** Social platforms let people easily communicate and exchange ideas or content.
- **Business and product marketing.** These platforms enable businesses to quickly publicize their products and services to a broad audience. Businesses can also use social media to maintain a following and test new markets. In some cases, the content created on social media is the product.
- **Audience building.** Social media helps entrepreneurs and artists build an audience for their work. In some cases, social media has eliminated the need for a distributor, because anyone can upload their content and transact business online. For example, an amateur musician can post a song on Facebook, get instant visibility among their network of friends, who in turn share it on their networks.

Challenges of social media

Social media can also pose challenges to individual users, in the following ways:

- **Mental health issues.** Overuse of social apps can result in burnout, social media addiction and other issues.
- **Polarization.** Individuals can end up in filter bubbles. They create the illusion of open discourse when the user is actually sequestered in an algorithmically generated online community.
- **Disinformation.** Polarized environments foster the spread of disinformation where the perpetrator's intent is to deceive others with false information.

Businesses face similar and unique social media challenges.

- **Offensive posts.** Conversations on intranets and enterprise collaboration tools can veer off into non-work-related subjects. When that happens, there is potential for co-workers to disagree or be offended. Controlling such conversations and filtering for offensive content can be difficult.
- **Security and retention.** Traditional data security and retention policies may not work with the features available in collaboration tools. This can raise security risks and compliance issues that companies must deal with.
- **Productivity concerns.** Social interaction, whether online or in person, is distracting and can affect employees' productivity.

Enterprise social media best practices

It is important for companies to have a social media strategy and establish social media goals. These help to build trust, educate their target audience and create brand awareness. They also enable real people to find and learn about a business.

Here are some social media social media best practices for companies to follow:

- Establish social media policies that set expectations for appropriate employee social behavior. These policies should also ensure social media posts do not expose the company to legal problems or public embarrassment. Guidelines should include

directives for when an employee must identify them self as a company representative and rules for what type of information can be shared.

- Focus on platforms geared to B2B marketing, such as Twitter and LinkedIn.
- Put in place an engaging, customer-centric strategy in social media campaigns. An example would be to use Twitter to field questions from customers.
- Include rich media, such as pictures and video, in content to make it more compelling and appealing to users.
- Use social media analytics tools to measure user engagement with content and to keep on top of trends.
- Use a conversational voice in posts that comes across as professional but not rigid.
- Shorten long form content to make it social friendly. Lists and audio and video snippets are examples.
- Embrace employees and customers talking positively about the organization and repost that content.
- Check in on analytics and management tools frequently, if not on a daily basis, as well as the social media accounts.

Types of social media:

The four main categories of social platforms are these:

1. **Social networks.** People use these networks to connect with one another and share information, thoughts and ideas. The focus of these networks is usually on the user. User profiles help participants identify other users with common interests or concerns. Facebook and LinkedIn are good examples.
2. **Media-sharing networks.** These networks focus is on content. For example, on YouTube, interaction is around videos that users create. Other media-sharing networks are TikTok and Instagram. Streaming platforms like Twitch are considered a subset of this category.
3. **Community-based networks.** The focus of this type of social network is in-depth discussion, much like a blog forum. Users leave prompts for discussion that spiral into

detailed comment threads. Communities often form around select topics. Reddit is an example of a community-based network.

4. **Review board networks.** With these networks, the focus is on a review, usually of a product or service. For example, on Yelp, users can write reviews on restaurants and endorse each other's reviews to boost visibility.

Examples of social media:

Here are some examples of popular web-based social media platforms:

- **Facebook** is a free social networking website where registered users create profiles, upload photos and video, send messages and keep in touch with friends, family and colleagues.
- **LinkedIn** is a social networking site designed for the business community. Registered members can create networks of people they know and trust professionally.
- **Pinterest** is a social curation website for sharing and categorizing images found online. The main focus of Pinterest is visual, though it does call for brief descriptions of images. Clicking on an image will take a user to the original source. For example, clicking on a picture of a pair of shoes might redirect a user to a purchasing site; an image of blueberry pancakes might redirect to the recipe.
- **Reddit** is a social news website and forum where site members curate and promote stories. The site is composed of hundreds of sub-communities called *subreddits*. Each subreddit has a specific topic, such as technology, politics or music. Reddit site members, also known as "redditors," submit content that members vote on. The goal is to elevate well-regarded stories to the top of the site's main thread page.
- **Twitter** is a free microblogging service for registered members to broadcast short posts called tweets. Twitter members can broadcast tweets and follow other active users' tweets using several platforms and devices.
- **Wikipedia** is a free, open content encyclopedia created through a collaborative community. Anyone registered on Wikipedia can create an article for publication; registration is not required to edit articles.

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
Sea Monkey	Mozilla Foundation
K-meleon	K-meleon

Architecture

There are a lot of web browser available in the market. All of them interpret and display information on the screen however their capabilities and structure varies depending upon implementation. But the most basic component that all web browser must exhibit are listed below:

- Controller/Dispatcher

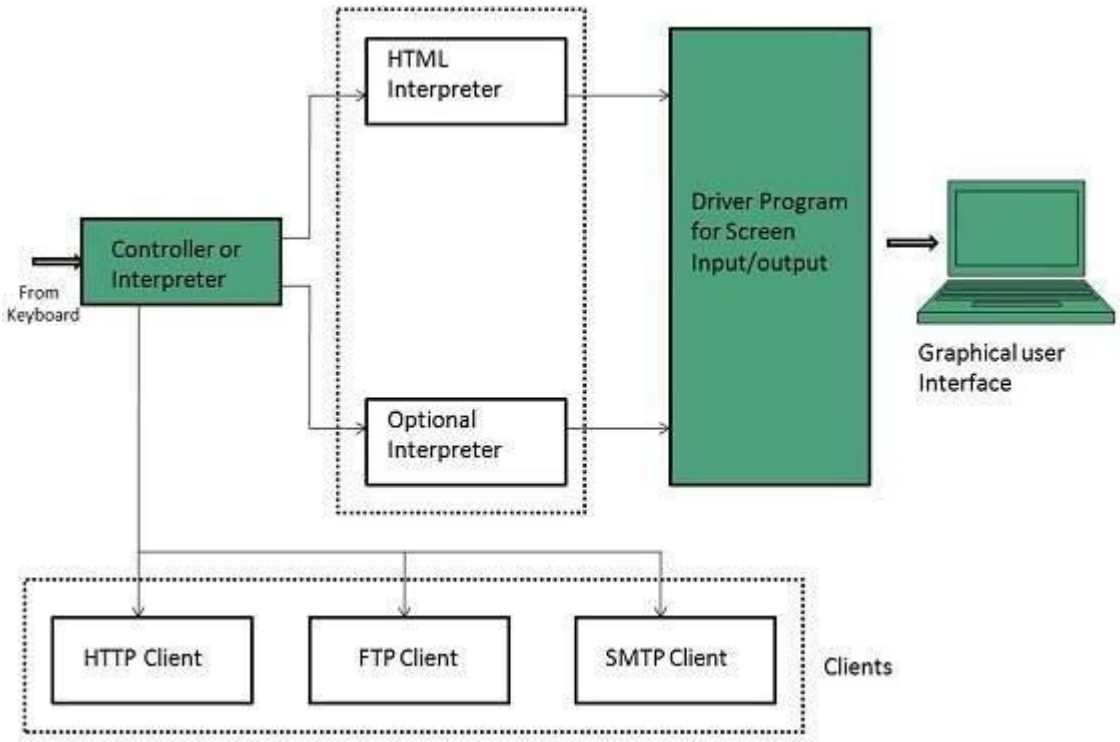
- Interpreter
- Client Programs

Controller works as a control unit in CPU. It takes input from the keyboard or mouse, interpret it and make other services to work on the basis of input it receives.

Interpreter receives the information from the controller and execute the instruction line by line. Some interpreter are mandatory while some are optional For example, HTML interpreter program is mandatory and java interpreter is optional.

Client Program describes the specific protocol that will be used to access a particular service. Following are the client programs tat are commonly used:

- HTTP
- SMTP
- FTP
- NNTP
- POP

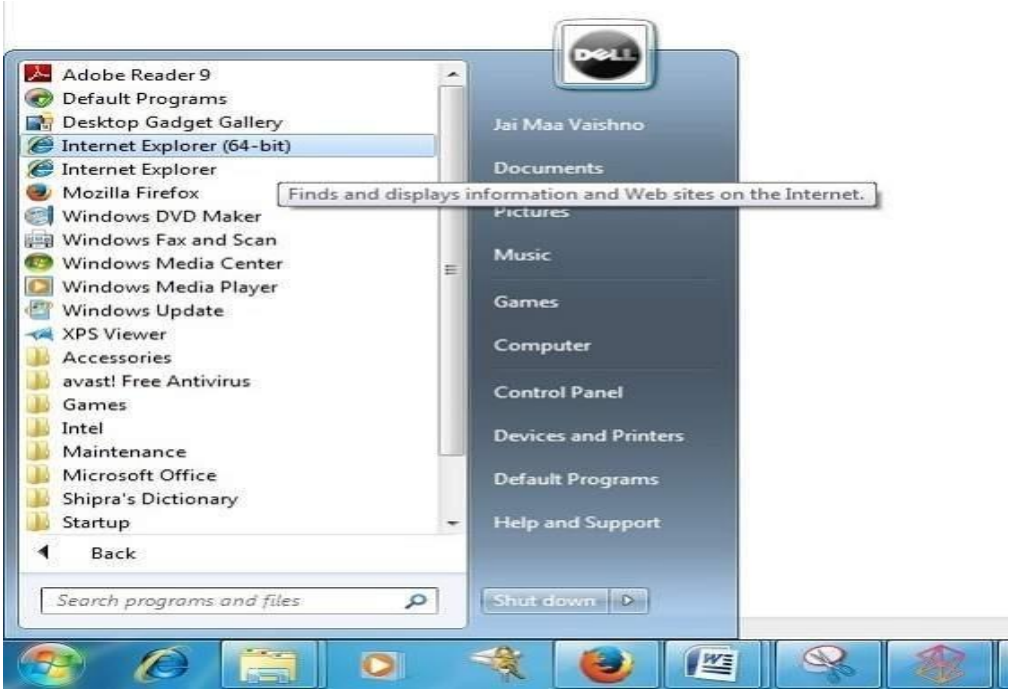


Starting Internet Explorer

Internet explorer is a web browser developed by Microsoft. It is installed by default with the windows operating system however, it can be downloaded and be upgraded.

To start internet explorer, follow the following steps:

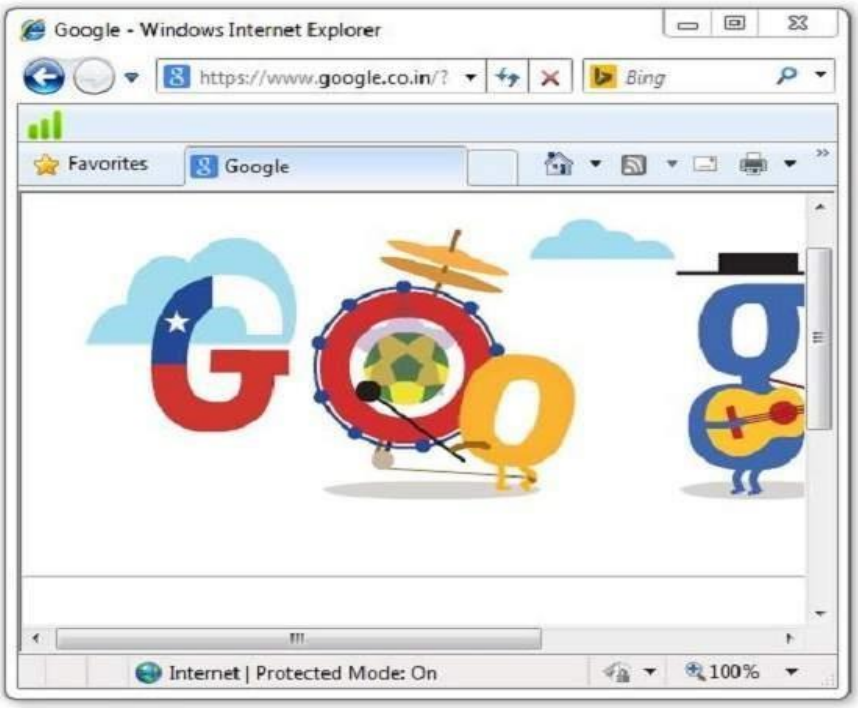
- Go to **Start** button and click **Internet Explorer.**



The **Internet Explorer** window will appear as shown in the following diagram:

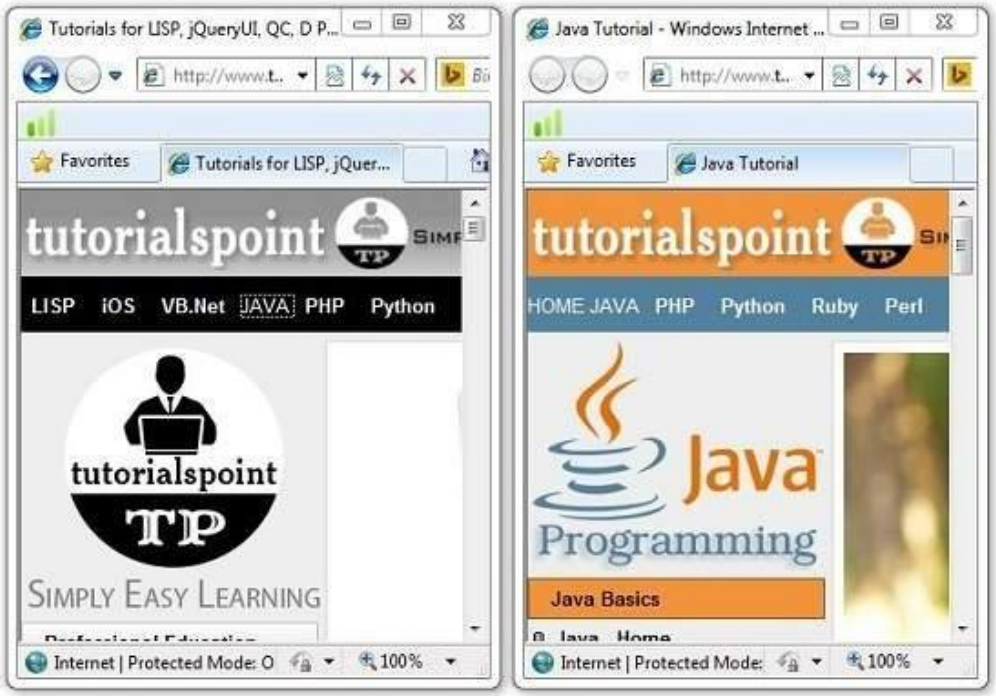
Accessing Web Page

Accessing web page is very simple. Just enter the **URL** in the address bar as shown the following diagram:



Navigation

A web page may contain **hyperlinks.** When we click on these links other web page is opened. These hyperlinks can be in form of text or image. When we take the mouse over an hyperlink, pointer change its shape to hand.



Key Points

- In case, you have accessed many web pages and willing to see the previous webpage then just click back button.
- You can open a new web page in the same tab, or different tab or in a new window.

Saving Webpage

You can save web page to use in future. In order to save a webpage, follow the steps given below:

- Click **File > Save As.** Save Webpage dialog box appears.
- Choose the location where you want to save your webpage from **save in:** list box. Then choose the folder where you want to save the webpage.
- Specify the file name in the **File name** box.
- Select the type from **Save as** type list box.
    - Webpage, complete
    - Web Archive
    - Webpage HTML only
    - Text File
- From the **encoding** list box, choose the character set which will be used with your webpage. By default, **Western European** is selected.
- Click **save** button and the webpage is saved.

Saving Web Elements

Web elements are the pictures, links etc. In order to save these elements follow the steps given below:

- **Right click** on the webpage element you want to save. Menu options will appear. These options may vary depending on the element you want to save.

Save Picture As: This option let you save the picture at specific location with its name. When you click this option, a dialog box is opened where you can sepcify its name and location.

Favourites

The Favourites option helps to save addresses of the webpages you visited oftenly. Hence you need not to remember long and complex address of websites you visit often.

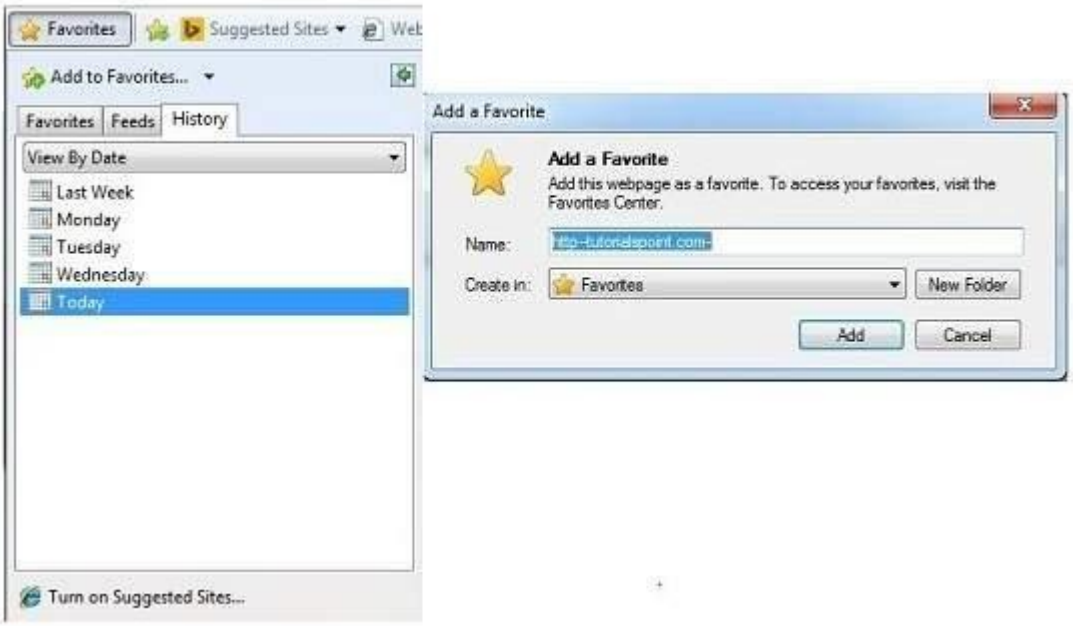
In order to open any webpage, you just need to double click on the webpage that you have marked from bookmarks list.

Adding a web page to your Favourites

In ordered to add website to your favourite list, follow the steps given below:

- Open webpage that you want to add to your favourite.
- Click on **favourite menu** and then click on **Add to Favourites** opton. **Addfavourites** dialog box appears.

You can also click **Favourites** button available in the toolbar. Favourites panel will open in the left corner of the internet explorer window. Click **add** button, **AddFavourites** dialog box will apppear.

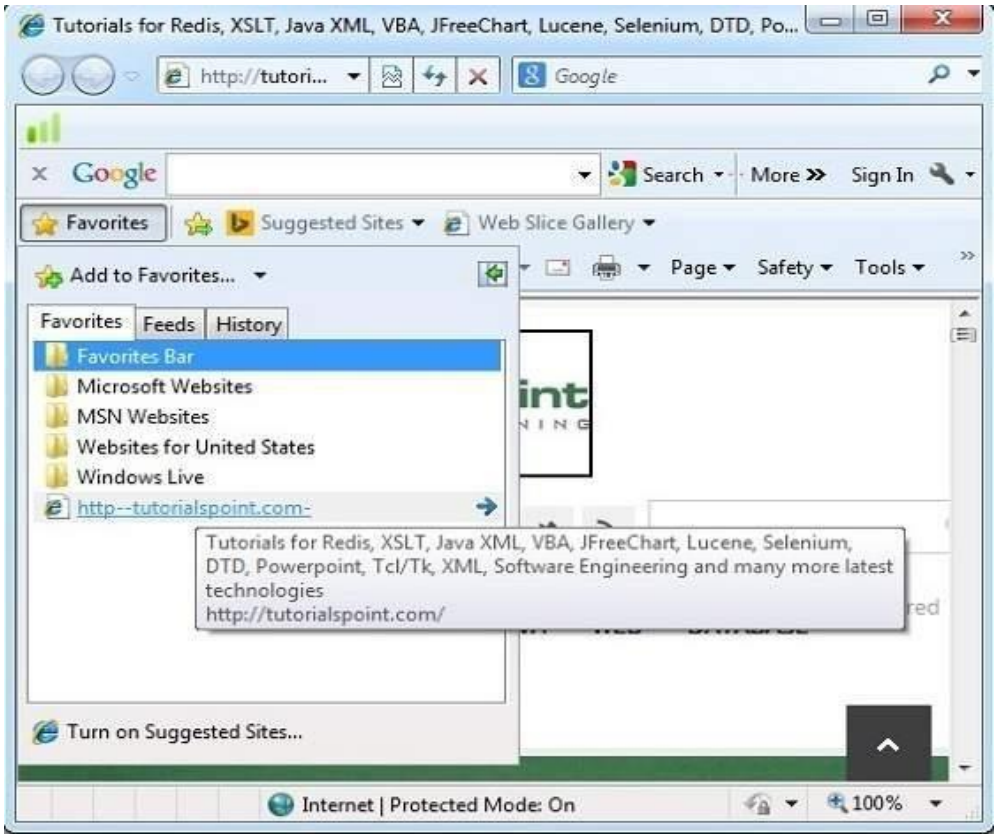


- In **AddFavourites** dialog box, the **Name:** text box will contains the name of the web page that you want to add to favourites.
- Click the **Create in** button, Favoutites folder will appear. Move to the folder where you want to store the favourites by clicking on the folder name.
- Now click **OK** button to save the favourites.

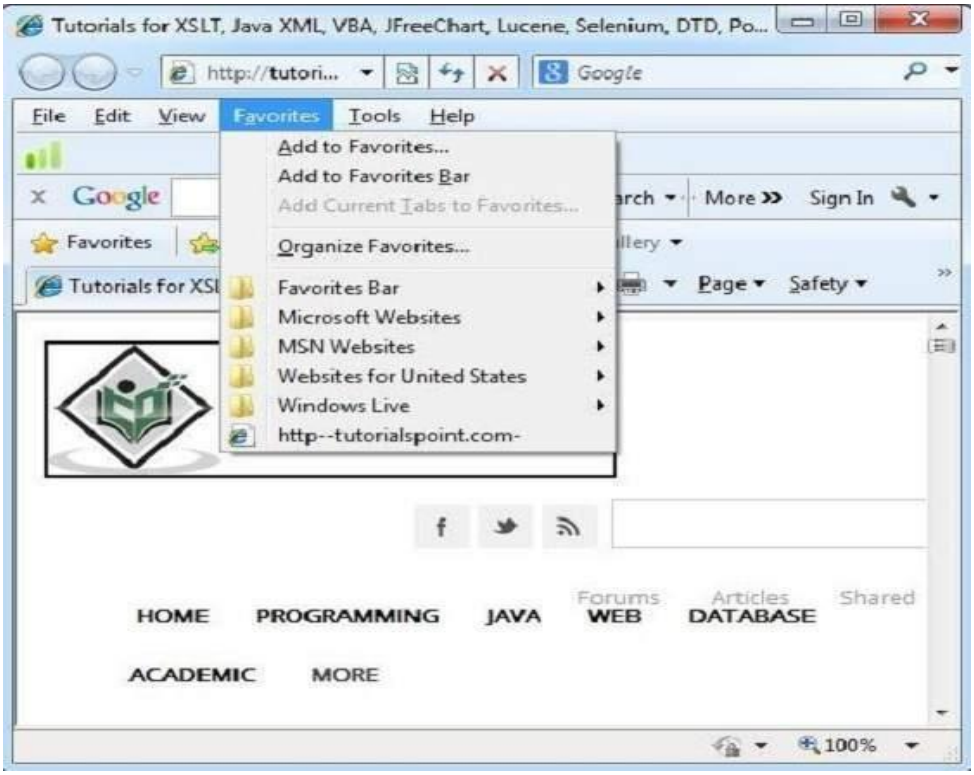
Opening Favourites

In order to open favourites, follow the steps given below:

- In the Favourite Panel, take the mouse over the site that you want to open. Now click on the address to open that site.



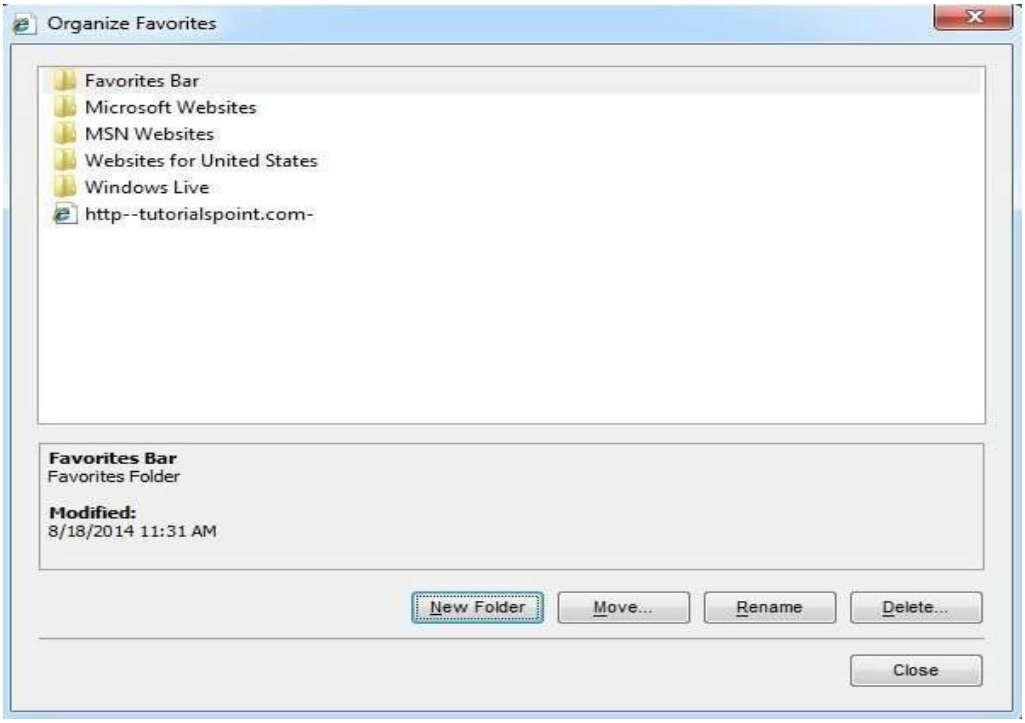
- Favourite can also be opened from the **Favourites** menu by selecting the appropriate one.



Organizing Favourites

Favourites can be organized by categorizing web pages, creating folder for each category and then storing web pages into them. In order to organize favourites, follow the steps given below:

- Click **Favourites menu > Organize Favourites.** Organize favourites dialog box will appears.
- In order to organize the webpages, drag the individual webpage to the respective folder. Similarly to delete a favourite, Click on **delete** button.



Web development comes with a huge set of rules and techniques every website developer should know about. If you want a website to look and function as you wish them to, you need to get familiar with web technologies that will help you achieve your goal.

Developing an app or a website typically comes down to knowing 3 main languages: JavaScript, CSS, and HTML. And while it sounds quite complicated, once you know what you are doing, understanding web technology and the way it works becomes significantly easier.

Fret not if it is not coming easily to you immediately. You may need more time, training, and patience to dive deeper into the subject, but you'll end up with a good understanding eventually.

We present you with an introduction to web technologies and the latest web technologies list hoping it will make things at least a bit easier for you. Now, let's take a look.

Web Technology

You have probably heard the term "web development technologies" before, but did you ever think about what it actually means?

Since computers can't communicate with each other the way people do, they require codes instead. Web technologies are the markup languages and multimedia packages computers use to communicate.

1. Browsers

Browsers request information and then they show us in the way we can understand. Think of them as the interpreters of the web. Here are the most popular ones:

Google Chrome – Currently, the most popular browser brought to you by Google

Safari – Apple's web browser

Firefox – Open-source browser supported by the Mozilla Foundation

Internet Explorer – Microsoft's browser

2. HTML & CSS

HTML is one of the first you should learn. Thanks to HTML, the web browsers know what to show once they receive the request. If you want to better understand how HTML works, you also need to know what CSS is.

CSS stands for Cascading Style Sheets and it describes how HTML elements are to be displayed on the screen. If you browse enough tutorials, you'll soon create CSS text effects, page transitions, image hover effects, and more.

If you're a complete beginner, this Essential HTML & CSS training by James Williamson will help you to quickly get started with these technologies.

3. Web Development Frameworks

Web development frameworks are a starting point of items that a developer can use to avoid doing the simple or mundane tasks, and instead get right to work.

Angular

Angular is one of the latest web technologies designed specifically for developing dynamic web applications. With this framework, you can easily create front-end based applications without needing to use other frameworks or plugins.

The features include well-made templates, MVC architecture, code generation, code splitting etc. All the expressions are like code snippets that enclosed within curly braces and do not use any loops or conditional statements.

If you would like to start using Angular or to just quickly evaluate if this framework would be the right solution for your projects, you can check out this 3-hour training, published in June 2019 by Justin Schwartzenberger, a Google Developer Expert. This course covers everything that's necessary to start using Angular, from basic architecture, work with DOM, data binding, routing, and components, to more advanced topics such as directives and pipes.

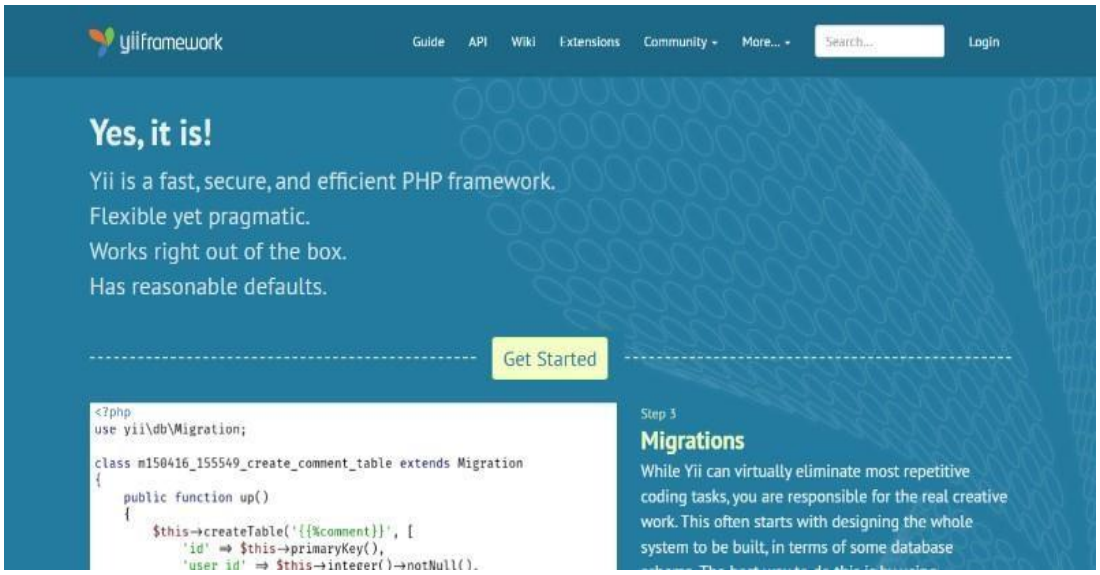
Ruby on Rails

Ruby on Rails is a server-side website technology that makes app development much easier and faster. The thing that really sets this framework apart is the reusability of the code as well as some other cool features that will help you get the job done in no time.

Popular websites written with Ruby include Basecamp, Ask.fm, GitHub, 500px, and many others.Here is everything you need to know about Ruby on Rails.

If you would are interested in a more in-depth training on Ruby on Rails framework, this 10-hour course by Kevin Skoglund, a senior Ruby developer, might be just the right resource to get started. It covers the complete learning cycle from the very fundamentals to more advanced topics such as Layouts, Partials, and View Helpers, giving quite a few practical tasks in parallel.

YII



Yii is an open-source web application development framework built in PHP5. It is performance-optimized and comes with a number of great tools for debugging and app testing. Another plus is that it is pretty simple and easy to use.

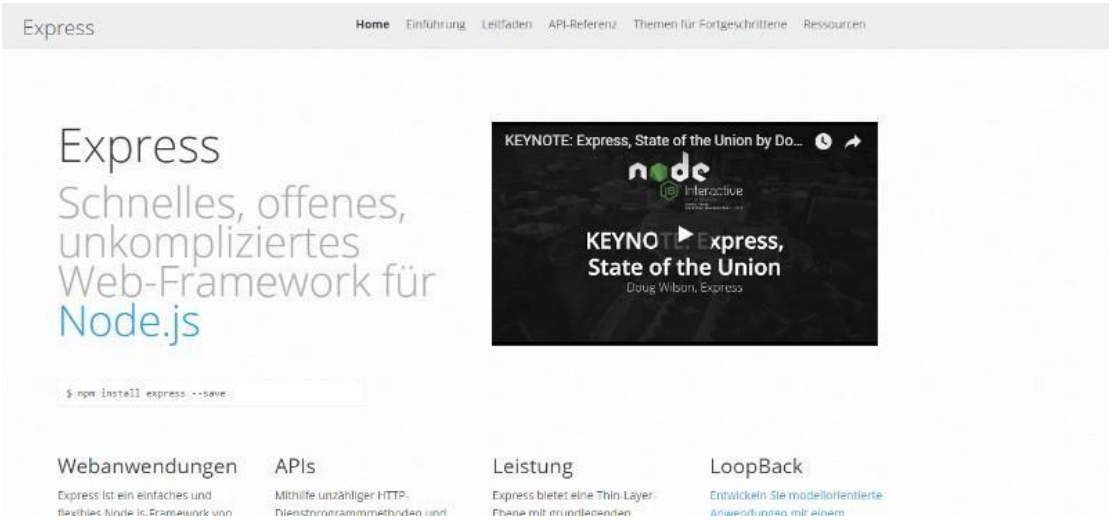
Meteor JS

Meteor JS is written in Node.js and it makes it possible for you to create real-time web applications for different platforms. The framework for creating simple websites for personal use really stands out with Meteor JS.

This is an open-source isomorphic JavaScript web framework which also means that the webpage loading time is significantly shorter. JavaScript stack also makes it possible to get the same results with fewer lines of code than usual.

This online video course gives an interesting practical example of combining MeteorJS and React to build a web app.

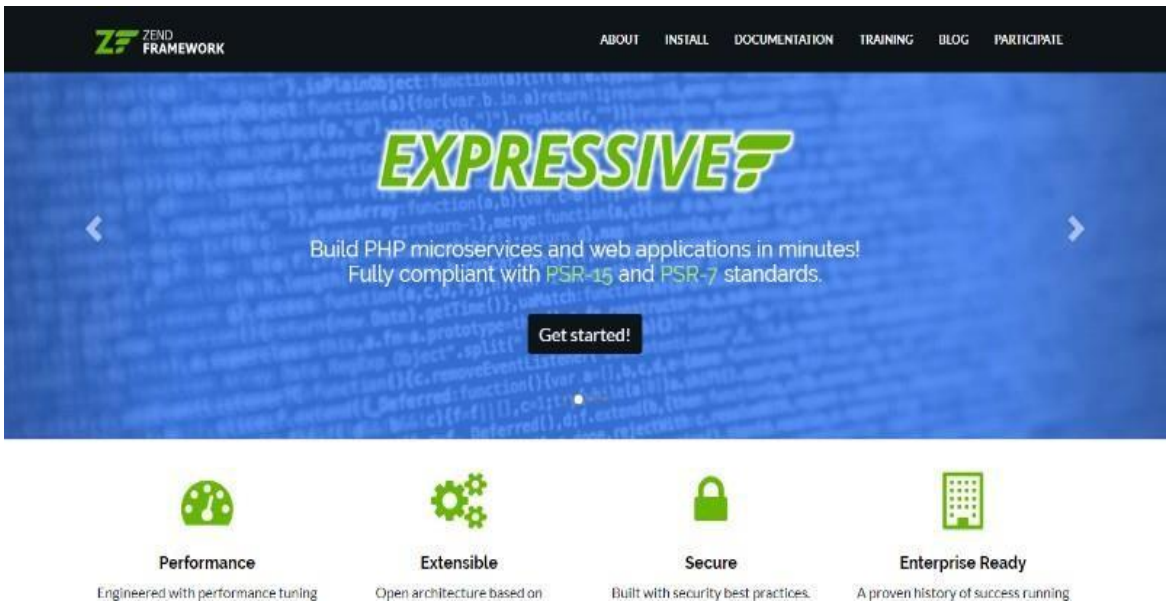
Express.js



Developed in Node.js, Express.js is a web app development network that is great for those who need to develop apps and APIs as fast as possible. A lot of great features are provided with the help of plugins.

This course provides a good insight into the advanced usage of Express.js in combination with MongoDB and Mongoose and shows different ways of deploying an Express app and running it in production.

Zend

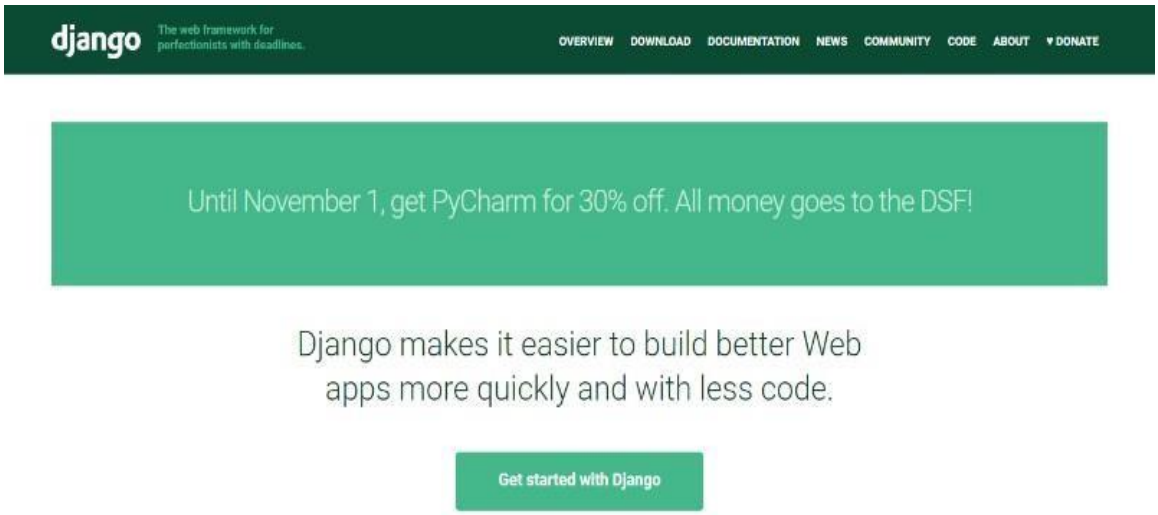


Zend is an open-source framework based on PHP, focused on building more secure and reliable web apps and services. It is one of the first enterprise-level MVC frameworks, which came before the current superhits such as Laravel or Symfony, and many popular PHP engines such as Magento were built in Zend.

Today Zend is still under active development, and even though it may be less popular than its opensource siblings, it is a great solution for a large-scale PHP app.

Watch this short video course where different PHP MVC frameworks are compared so that you could make a choice on your own.

Django

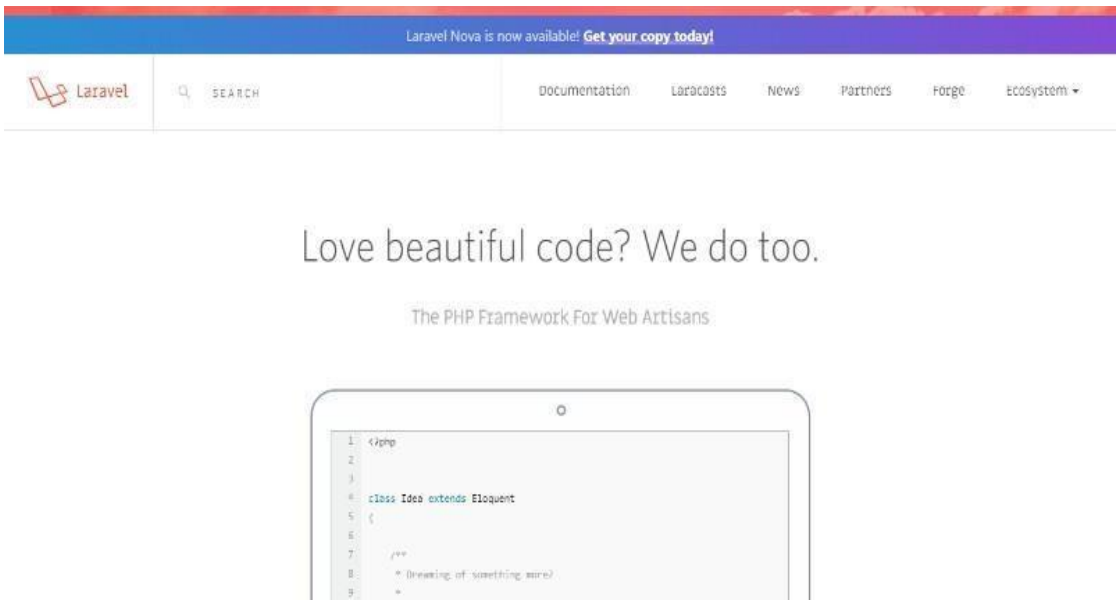


Django is one of the most popular frameworks written in Python and follows MVC architecture. It makes the app development process much easier thanks to its simplicity.

Django simplifies using Python a lot and provides multiple tools that make a web app developer's life easier – e.g. an ORM, Models, Django admin, templates, etc. This 1.5-hour video course can help any developer, even a beginner, to start developing Python/Django apps in a couple of days.

Check out more popular Python frameworks.

Laravel



Laravel is a PHP development framework ideal for small websites. It comes with a number of useful features including the MVC support, object-oriented libraries, Artisan, authorization technique, database migration, etc. Currently, it is one of the most community-supported and community-developed frameworks, and given that PHP has one of the largest communities out there, Laravel is a great tool powering both small websites and large-scale B2B web apps managing millions of transactions daily.

To get started with Laravel in less than 3 hours, watch this video course by Bernando Pineda, a senior DevOps, and Engineer with 15+ years of software development experience.
It is one of our favorite PHP frameworks.

4. Programming Languages

As we explained before, since computers don't use languages that are anything like human languages, they need a different way to communicate. Here are some of the most popular programming languages:

Javascript – used by all web browsers, Meteor, and lots of other frameworks
CoffeeScript – a "dialect" of JavaScript. It is viewed as simpler but it converts back into JavaScript

Python – used by the Django framework as well as in the majority of mathematical calculations

Ruby – used by the Ruby on Rails framework

PHP – used by WordPress to create those WYSIWYG editors that everyone is using now. It's also used by Facebook, Wikipedia, and other major sites

Go – newer language, built for speed

Swift – Apple's newest programming language

Java – used by Android and a lot of desktop applications.
So let's talk about the most popular ones in a bit more detail.

JavaScript

According to StackOverflow's annual survey, JavaScript is the most popular programming language with 62.5% of respondents claiming to use it.

It is one of the core web technologies and if you want to learn more about it, you can start with this essential training that covers all the basics, working with functions and objects, interacting with DOM, etc. This course is recent – from April 2019 – Javascript evolves quickly, so it makes sure you leverage the newest language "perks" as you learn.

Ruby

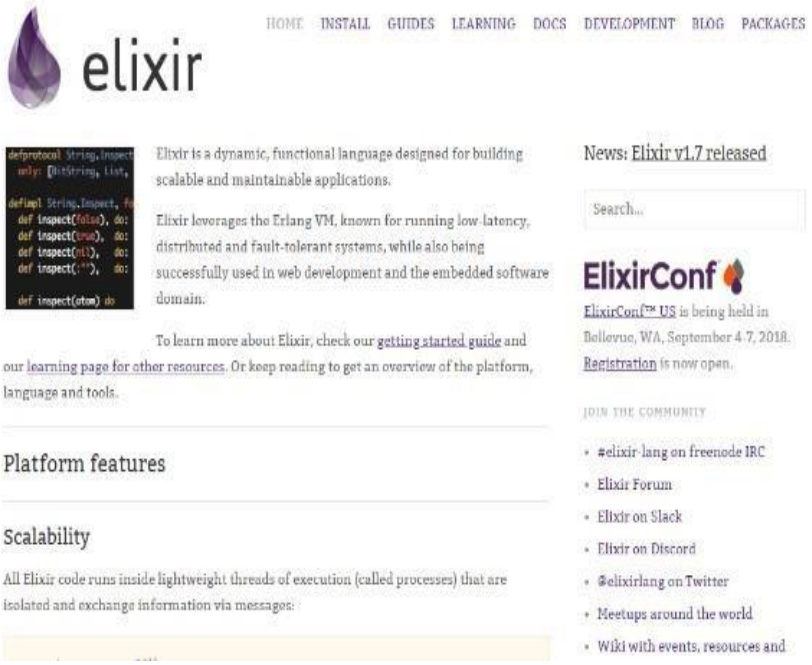
The developers love Ruby – and for all the right reasons. Designed to be user-friendly and really easy to use, it's no wonder that this programming language is often called "a programmer's best friend."

What you can expect from Ruby is a shorter, readable code. Unfortunately, that sometimes means lower efficiency compared to other programming languages – but it also means higher productivity.

If you are a beginner in the web development world, Ruby would be a great choice for the first programming language to learn. A well-written Ruby code can be almost as readable as the sentence in plain English language.

But the real reason most people use Ruby is its popular framework — Ruby on Rails which we mentioned earlier in the text. The great productivity achieved with Rails makes it a common choice for startups who aim for a running start.

Elixir



Elixir appeared back in 2011 and gained popularity almost immediately. It was inspired by Erlang, a language developed back in the '80s by Ericsson. Elixir's author José Valim himself said that he loved Erlang, but also noticed some things that could use a bit of improvement.

Scala

Scala stands for Scalable Language, and is one of the many attempts to "rewrite Java" and it is compiled to run on the Java Virtual Machine (JVM). It is safe to say this programming

language turned out to be quite a success taking into consideration that companies like LinkedIn, Twitter, and The Guardian use it in their codebases. Scala is known to be a complex language but also a language worth learning.

5. Protocols

The instructions for how to pass information back and forth between computers and devices are commonly known as protocols.

HTTP

Thanks to this protocol, each website can get to the browser. The protocol requests the website from Google's server and then receives a response with the HTML, CSS, and JavaScript of the website.

DDP

Uses WebSockets to create a consistent connection between the client and the server. As a result of that, you get website updates in real-time without having to refresh the browser.

REST

Used mostly for API's, this protocol has standard methods like GET, POST, and PUT that let information be exchanged between applications.

6. API

An API (application programming interface) allows other developers to use some of the app's functionality without sharing the code.

The endpoints are exposed by the developers while the API can control access with an API key. Examples of well-made APIs are those created by Facebook, Twitter, and Google for their web services.

7. Data formats

Data is stored in a structure called a data format.

JSON – JavaScript Object Notation is a syntax for storing and exchanging data (just like XML). It is currently becoming the most popular data format out there.

XML – Predominantly used by Microsoft systems, it used to be the most popular data format

CSV – is data formatted by commas; for example Excel data

Email*

8. Client (or Client-side)

Each user of an application is called a client. Clients can be computers, mobile devices, tablets etc. Usually, multiple clients are interacting with the same app stored on a server.

9. Server (or Server-side)

The application code is usually stored on the server. The clients make requests to the servers. The servers then respond to those requests after gathering the requested information.

Ending thoughts on the latest web technologies

In order to stay up to date with the latest web technologies, one has to learn new things all the time. Web technologies are being improved and updated all the time and every web development team should take advantage of that whenever possible.

New web technologies change the entire web development process and it can be hard sometimes to understand all of them in the right way. Luckily, with the right internet technology tutorial, you should be able to learn more about them in no time.

If you enjoyed reading this article on web technologies, you should check out this one about web application development.

We also wrote about a few related subjects like web development trends and web application testing.

Design in the Browser

Design in the browser is basically a concept using HTML and CSS as your primary design tools.

Basically the code is writing right from the scratch during each phase of the project. From the clients brief on to the first design draft, to an rudimentary prototype to a finished product. Everything (or nearly everything) takes place in the browser.

Rather than spending hours designing pixel-perfect design drafts in Photoshop, designing in the browser allows you to jump directly into the text editor and start shaping your code.

TOOLS FOR DESIGNING IN THE BROWSER:

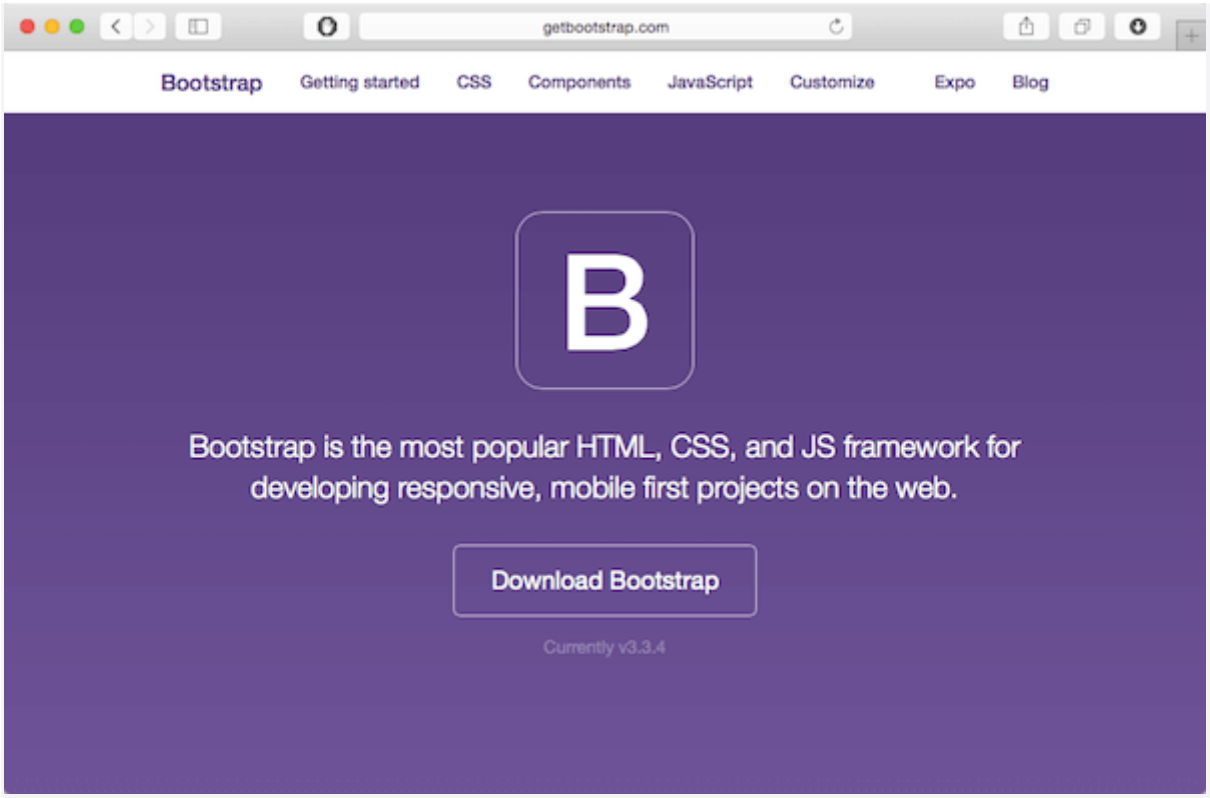
Convinced of the "design in the browser" concept? Want to get right into it? Perfect, here are a couple of great tools which are making your life easier when designing in the browser.

Get a programming friendly editor

This might sound a bit strange. But designing in the browser basically means working a lot on the code of your site. The editor will be your friend and go-to tool in many cases. Choose your editor wisely.

Bootstrap – your front-end framework

Bootstrap is probably the best known front-end prototyping framework available. Originally designed by Twitter, it's now available to everyone for free. Packed with some great functionalities, it supports typography, forms, buttons and some great JavaScript options.



http://getbootstrap.com/

PS: Foundation – similar to Bootstrap – is another, yet great front-end framework which is worth a try.

Style guide

Next, it's about the style guide. It's super important to keep your design and style elements organized. With a style guide in place, design changes are super easy as they will come. And trust me: they will come □

http://styletil.es/

Chrome Developer Tools

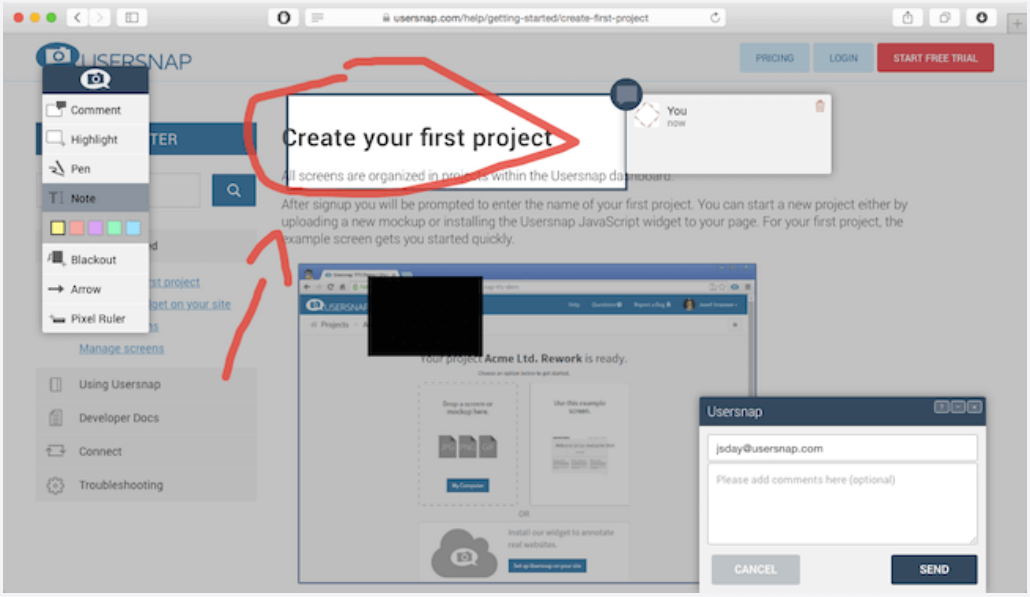
After you created your first prototype in your browser, it's time to review, test and tweak. The best tools therefore are available for free in every Chrome browser (or Firefox if you prefer).

With a right click on your site you can start the Chrome developer tools which offer you a broad range of features. You can play around with your code, move styles, edit content and much more.

Keep track & manage tasks with Usersnap

Since designing in the browser is quite an agile approach, you might wonder how you keep track of change requests, bugs, ideas, etc. I therefore recommend Usersnap, which does a great job making collaboration on website or web app projects much easier.

By adding the Usersnap widget to your prototype, every single piece of comment and feedback will be stored in your project overview where you can discuss feedback and track bugs.



Design skills still needed.

Designing in the browser still requires great design skills. No tool in the world will make up for your lack in design know how. However, the browser becomes more and more your design and development environment. And because most of the design process takes place in the browser, doesn't mean you should give up Photoshop.

Web Hosting and Publishing

Several web hosting options are available to campus, with different levels of complexity and configurability. Below is a collection of options available.

Creating a website to showcase your research doesn't have to be complicated. We've collected the top free (or affordable) resources across campus that are at your disposal.

News, blogs, and latest updates: WordPress or WebTools Blogs

WordPress is one of the most popular content management systems for blogs and news sites.

Publish.Illinois.edu is a multi-site WordPress installation maintained by Technology Services which offers semi-customizable templates to choose from. All the patching and maintenance is taken care of for you; all you need to do is provide the content.

cPanel Self Service Web Hosting can host WordPress sites and other site types as well. If you're looking at WordPress but find Publish's restrictions too confining, you can install your own WordPress instance on cPanel, and you'll also have responsibility for patching and maintenance.

The WebTools Blog service allows University members to create blogs or lightweight websites. These can then easily integrate with other WebTools such as calendars, newsletters, forms, surveys, and more.

Collaborative spaces with collective editing: Illinois Wiki, Google Docs, Office Online

Public-facing websites, news, and blogs are helpful when you have results to announce. However, when you want an internal collaboration space for your workgroup as you work on your project, you may want a solution more like a wiki or shared document spaces.

Illinois Wiki is a vended wiki solution called Atlassian Confluence, and is used extensively for class collaboration and help documents. Wiki spaces provide the large-group shared

editing and version control capabilities of a system like Google Docs, but also offer web-like navigation and organization rather than storage-folder-like organization.

Google Apps @ Illinois include Google Drive for shared storage and Google Docs for collective editing of files. If you don't need to present a web-like navigation structure, Google Docs is one of the most convenient options for large group editing of shared documents with stored version history.

Microsoft's Office Online tools allow you to collectively edit Microsoft Office document formats that are stored in a variety of locations, including U of I Box, OneDrive, and more. The experience is similar but not identical to editing on a desktop application; if version history and comments are important to you, you'll find the desktop version valuable.

Scholarly publishing on the Web: IDEALS, IDB, Omeka, Scalar, PWW

One key difference between "web hosting" and "web publishing" is how much the end result resembles a book or journal. Most "web hosting" offerings are for sites that are regularly updated and change frequently.

However, sometimes you want the digital equivalent of publishing a book: a fixed object that will be available in the same form at the same place for a long time, without regular edits (and sometimes without any changes at all). This is where "web publishing" comes into play.

Many of the resources for web publishing are offered through the University Library's Scholarly Communication and Publishing unit.

IDEALS is a digital repository for research and scholarship, including published and unpublished papers as well as other content types.

The Illinois Data Bank is a public access repository for publishing research data from the University of Illinois at Urbana-Champaign. The IDB can host data sets that supplement papers stored in IDEALS or other journals around the world.

Publishing Without Walls is a digital press that produces book-scale digital works that can include multimedia components. It is a digital scholarly publishing initiative that is scholar-driven, openly accessible, scalable, and sustainable.

If you'd like to produce the digital equivalent of a museum exhibit, with objects on display paired with descriptions and annotations of the resources, you may be interested in either the Library's offering of access to Omeka.net or installing your own Omeka copy on cPanel.

The Library also offers access to the University of Southern California's Scalar installation, for long-form narratives with interconnected multimedia experiences.

Small Questions

S. No	Questions	LOCF Mapping
1.	Define social media.	K1
2.	Name any four types of social media platforms.	K1
3.	What is a web development framework?	K1
4.	List any three programming languages used for web development.	K1
5.	What is an API?	K1

Big Questions

S. No	Questions	LOCF Mapping
1.	Discuss the business applications, benefits, and challenges of social media.	K2, K3
2.	Explain the different types of social media with examples.	K2
3.	Describe the various web development frameworks and their features.	K2
4.	What is the concept of "Design in the Browser"? Discuss the tools used for it.	K2, K3
5.	Explain the various options available for web hosting and publishing.	K2

UNIT IV: Search Engines and Networks

Content: Search Engines – Introduction, Market Share, Major Search Engines, Directories; Search Provider Relationships; Components of Search Engine; Ranking Factors; Search Engine Spam; Intranet and Extranet; Design and Evaluation of Search Engines.

UNIT -4

INTRODUCTION TO SEARCH ENGINES

As the Internet started to grow and became an integral part of day-to-day work, it became almost impossible for a user to fetch the exact or relevant information from such a huge web. This is the main reason why 'Search Engines' were developed. Search engines became so popular that now more than 80% of web-site visitors come from them. What exactly is a Search Engine? According to web opedia, a "Search Engine" is a program that searches documents for specified keywords and returns a list of the documents where the keywords were found".

For Example, if you want to know about the Automobile market in Canada, you will type keywords like automotive market, automobiles in Canada, automobile manufacturers in Canada etc... Once you click on the search button, you'll get the best relevant data related to those keywords.

On the eve of Google's initial public offering, new surveys and traffic data confirm that search engines have become an essential and popular way for people to find information online. A nationwide phone survey of 1,399 Internet users between May 14 and June 17 by the Pew Internet & American Life Project shows:

• 84% of internet users have used search engines. On any given day online, more than half of those using the Internet use search engines. More than two-thirds of Internet users say they use search engines at least a couple of times per week.

• The use of search engines usually ranks only second to email use as the most popular activity online. During periods when major news stories are breaking, the act of getting news online usually surpasses the use of search engines.

• There is a substantial payoff as search engines improve and people become more adept at using them. Some 87% of search engine users say they find the information they want most of the time when they use search engines.

The convenience and effectiveness of the search experience solidifies its appeal. Some 44% say that most times they search they are looking for vital information they absolutely need.

COM Score Networks tracking of Internet use shows that among the top 25 search engines:

• Americans conducted 6.7 billion total searches in December.

• 44% of those searches were done from home computers, 49% were done from work computers, and 7% were done at university-based computers.

• The average Internet user performed 33 searches in June.

• The average visit to a search engine resulted in 4.4 searches.

• The average visitor scrolled through 1.8 result pages during a typical search.

• In June, the average user spent 41 minutes at search engine sites.

• COM Score estimates that 40-45 percent of searches include 1sponsored results.

• Approximately 7 percent of searches in March included a local modifier, such as city and state names, phone numbers or the words “map” or “directions.”

• The percentage of searches that occurred through browser toolbars in June was 7%.

Search engine market share

Four times voted as Most Outstanding Search Engine, Google is an undisputed market leader of the search engine industry. Google is a crawler based search engine, which is known for

providing both comprehensive coverage of web pages and most relevant information. It attracts the largest number of searches and the number goes up to 250 million searches everyday.

Yahoo! is the second largest player in the industry with 28% of market share. Yahoo! started as a human based directory but turned into a Crawler based search engine in 2002. Till early 2004, it was powered by Google but after that they started to use their own technology.

Yahoo stands next to Google in terms of number of searches per day. It is owned by Yahoo and attracts more than 167 million searches a day. Yahoo was the first search engine to come up with a PPC program. AskJeeves initially gained fame in 1998 and 1999 as being the "natural language" search engine that let you search by asking questions and responded with what seemed to be the right answer to everything. When launched, it was run by around 100 editors who monitored search logs. Today, however, AskJeeves depends on crawler-based technology to provide results to its users.

Major Search Engines and Directories

Google: Right from the establishment in 1999, until today, Google is still the most popular search engine on the internet. Since its beta release, it has had phrase searching for NOT, it did not add an OR operation until Oct. 2000. In Dec. 2000, it added title searching. In June 2000 it announced a database of over 560 million pages, which grew to 4 billion by February 2004. Its biggest strength is its size and scope. Google includes PDF, DOC, PS, Image and many other file type indexing. It also has additional databases in the form of Google Groups, News, Directory, etc.

Yahoo!: Yahoo! is one of the best known and most popular internet portals. Originally just a subject directory, now Yahoo! is a search engine, directory and portal. It includes cached copies of pages and also includes links to the Yahoo! directory. It supports full Boolean searching, but it lacks in providing some advanced search features such as truncation. It indexes the first 500KB of a web page and link searches require inclusion of http://

Bing: Bing Search by Microsoft is the search engine for the MSN portal site. For years it had used databases from other vendors including Inktomi, LookSmart, and Direct Hit. As of February 1, 2005, it began using its own, unique database including separate News, Images, and Local databases along with links into Microsoft's Encarta Encyclopedia content. Its large and unique database, query building Search Builder and Boolean searching, cached copies of

web pages including date cached and automatic local search options are its strengths. However, limited advanced features, inconsistent availability of truncation are its weaknesses.

Ask: Debuting in spring 2001 and re-launching in April 2002, this new search engine has built its own database and offers some unique search features. It lacks full Boolean and other advanced search features, but it has more recently expanded and improved its search capabilities and added an advanced search. While Teoma results can show up in three separate sections, there is only the one single database of indexed Web pages. It may also include paid ad results (from Google's AdWords database) under the heading of 'Sponsored Links.' No additional databases or portal features are directly available. AskJeeves switched to Teoma instead of Direct Hit in January 2002 for the search engine results after its question and answer matches. Identifying Metasites and Refine feature to focus on web communities are the strengths while a smaller database, no direct URL submissions and no cached copies of pages are its weaknesses.

Directories

A Web Directory is a web search tool compiled manually by human editors. Once websites are submitted with information such as a title and description, they are assessed by an editor and, if deemed suitable for addition, will be listed under one or more subject categories. Users can search across a directory using keywords or phrases, or browse through the subject hierarchy. Best examples of a directory are Yahoo and the Open Directory Project.

The major difference between search engine and directory is the human factor. A web site search directory indexes a web site based on an independent description of a site. While directories perform many of the same functions of a web page search engine, although their indexing format is different. The main difference is that directories do not spider your site to gather information about it. Instead they rely on a few text entries, typically a site title, domain name, and description to determine which keywords describe your site. While sites in the search engines are scanned and resulted by program (crawler), they are edited manually in directories. Directories contain groups of websites according to theme or industry i.e. automobile related sites are placed in one sub-directory, sports sites are placed into the other sub-directory and so on. Directories do effectively help organize thousands of web sites together. A directory contained inside another directory is called a subdirectory of that directory. Together, the directories form a hierarchy, or tree structure.

There are 5 types of directories namely Human Edited, User Categorized, User Classified, Independently Classified and Pay Per Click (PPC). DMOZ and Yahoo! are the largest directories in the world today.

Search Provider Relationship

There are thousands of search engines available on the internet. Since it's not possible for all of them to create, maintain and update their own database, most display results from major search engines on their SERP (search engine results page).

It is not necessary that all primary and secondary results be provided by a single search engine. Different search engines can provide different results to other engines. Directories can also be used by a third party. The supplier and receiver relationship is demonstrated between many different search engines. These relationships are very important to understand if you want top rankings for your site.

Now let's check out the relationship between the top 10 search engines and top 2 directories i.e. which search engine is a supplier and which is the receiver.

Search Engines:

1. Google:

• Google's main search results are provided solely from Google's search technology, offering results from no other engine or source.

• The Google Directory is comprised of listings from The Open Directory Project (ODP, DMOZ).

2. Yahoo!:

• Paid and free submissions.

• Provides sponsored results from paid advertising sources.

3. Bing:

• Bing provides sponsored results from paid advertising sources.

• Paid and free submissions.

4. AOL:

• AOL results for "Recommended Sites" are listings that have been hand picked by AOL editors.

• AOL "Matching Sites" are supplied by Google results. The results in AOL may not always match the results in Google as Google often updates their database more frequently.

• AOL directory listings are provided by the ODP.

5. Alta Vista:

• Alta Vista receives sponsored listings from their own advertisers.

• Alta Vista will use results from their own database for the main search results.

• Alta Vista obtains its directory results from LookSmart.

6. HotBot:

• HotBot results contain three categories: Top 10 Results, Directory Results & General Web Results.

• Top 10 results include popular sites and searches.

• Directory results are hand-picked by human editors. • Web Results are provided by Inktomi.

7. IWon:

• IWon Spotlight results are comprised of web pages found within IWon or web sites that IWon has a direct business partnership with.

• IWon Sponsored Listings are provided by a variety of paid advertisements through third party pay for performance listings including Google, AdWords and Yahoo.

8. Lycos:

• Lycos provides directory results from The Open Directory Project.

• Lycos provides sponsored listings from Yahoo.

• Lycos provides Web Results from Fast and from the Lycos network.

9. Netscape:

• Netscape's sponsored links are provided by Google AdWords.

• Netscape's matching results include sites that are handpicked by ODP editors mixed with results powered by Google.

10. AllTheWeb:

• AllTheWeb crawls and index ODP results.

• AllTheWeb powers the main search results in Lycos.

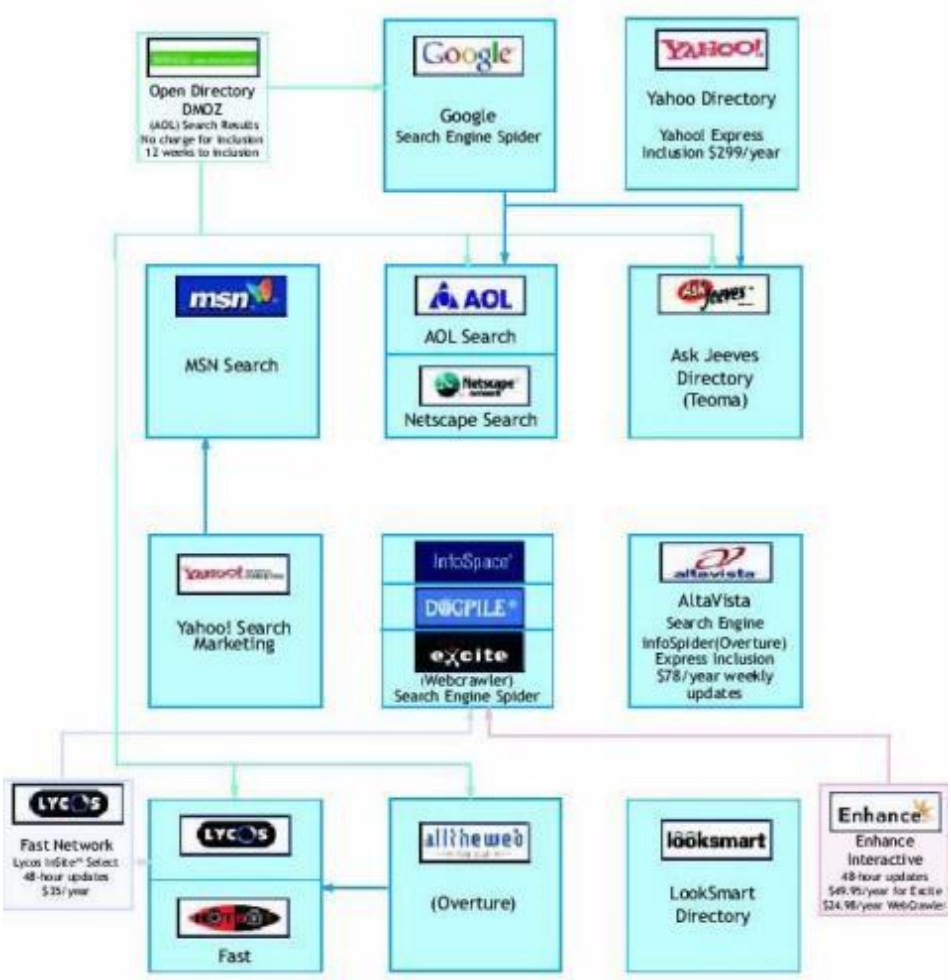
• AllTheWeb provides results from Lycos.

• AllTheWeb also powers the Lycos advanced search feature, the FTP search feature and their MP3 specialty engine.

Directories:

1. Dmoz: Directory listings are provided to AOL, Google, Lycos and Netscape and many other web sites, directories & portals.

2. Yahoo!: Yahoo! Directory listings are supplied by Yahoo! editors and require a fee for commercial sites. Yahoo directory results are provided to Alta Vista.



Components of Search Engine

How Search Engines Rank Pages

Broadly search engines are divided into 2 main categories:

a. Crawler based search engines

b. Human powered directories

Web crawler is a program, developed to scan the web page. The Crawler scans the entire page, indexes it and lists it on the search engine. It evaluates any particular web page based on several different factors such as keywords, table, page titles, body copy etc. Since listings are done automatically, it can change if you change some content of the website.

Manual listing is done in case of 'Human Powered Directories'. Human editors are responsible for listing any site in the directory. Webmasters need to submit a short

description to the directory for the entire site and a search looks for matches only in the description submitted. Listings are not affected if you change some content in your web site. Listing in directories and search engines are totally different and hence parameters for listing are different in both cases. With either listing, it's still necessary to create an informative and content rich site to attract more visitors.

Any crawler based search engine is made up of 3 basic components.

a. Crawler or Spider

b. Index

c. Search engine software

All these components work one after one and list the page on search engines. Search engines find websites in 2 ways:

1. By accepting listings sent by webmasters

2. By crawlers that roam the internet storing links and information about each page they visit. Once the site is found by the search engine, crawlers scan the entire site. While scanning, the crawler visits the web page, reads it and then follows link to other pages within the site. Major search engines like Google, Yahoo and MSN use multiple search engines simultaneously.

Google uses 4 spiders which crawl over 100 pages per second and generating around 600KBs of data each second.

Then index program starts after the crawler. Once a webpage is crawled, it is necessary to transfer them to the database. The index contains a copy of each web pages scanned by the crawler. If the webpage is changed, the index is updated with the new information. It is very important that your pages are added to the index. Until and unless it is indexed, it is not available to those searching with the search engines.

The search engine software performs a task of relevant listings. It searches the entire database i.e. indexed pages and matches it with the search. It then ranks and lists the most relevant matches. These listings are done on how the search engine software is programmed. It delivers listings according to what it believes the most relevant content is!

There are many more factors on which search engines rank a page but we will look at them in detail later.

Broadly, it depends on **On-page factors** and **Off-page factors**. On-page factors include keyword targeting, HTML tags, Content, Anchor Text and URL while Off-page factors include Link Building, Link Popularity and Anchor Text.

Though these terms are explained later, right now let's see the strategies in which search engines opt to list a page. Crawler based search engines list the sites without any human interference. This means it ranks a page based on what it thinks the most relevant page is! There are few parameters on which crawlers check whether the site is relevant to the search query or not. This program is called Search Engine Algorithm. No one knows the exact algorithm of any search engine but studies and research has proven that there are few factors, which are common in most search engine algorithms.

Location of keywords: Once keywords are finalized the main task is 'placement of keywords'. The search engine algorithm mainly revolves around the location of keywords. The keywords can be placed in HTML tags, content, headline or in the first few paragraphs. The importance varies according to location. Keywords placed in the headline or first few paragraphs are more important than other locations in web site. If keywords are placed from the beginning, search engines assume that the page is more relevant to that particular theme.

Frequency: Though it's very important to place keywords in the most visible parts of the web page, it is important to limit the number of keywords. This is called frequency. Search engines also measure frequency of keywords while ranking a page. Search engine analyses how often keywords appear in relation to other words in a web page therefore websites with a larger number of keywords are considered to be more relevant than others.

Added features in Location and Frequency: Location and frequency are just the basics of search engine algorithm. Once search engines discovered that anyone can play around it and can successfully rank their pages, they increased the algorithm complexity. Different search engines now index different number of web pages. Some index more and some less. Since some index more pages than others, no search engine has the exact same collection of web pages to search through.

Once webmasters came to know about the frequency, they cracked the algorithm by using too many keywords in a page, just to get higher rankings. Hence, search engines started to

penalize such sites with too many of the same keywords. Search engines termed it as ‘spamming’. It became very necessary for SEO companies to keep the frequency more than others but less than spamming. Search engines watch for common spamming methods in a variety of ways, including following up on complaints from their users.

Off-page factors: Above mentioned are some on-page factors. Now let’s look at some common off page factors. Crawler-based search engines have plenty of experience now with webmasters who constantly rewrite their web pages in an attempt to gain better rankings. Some sophisticated webmasters may even go to great lengths to "reverse engineer" the location/frequency systems used by a particular search engine. Because of this, all major search engines now make use of "off the page" ranking criteria.

Off the page factors are those that a webmasters cannot easily influence. Chief among these is link analysis. By analyzing how pages link to each other, a search engine can both determine what a page is about and whether that page is deemed to be "important" and thus deserving of a ranking boost. In addition, sophisticated techniques are used to screen out attempts by webmasters to build "artificial" links designed to boost their rankings.

• **Link analysis:** Web-based search engines have introduced one dramatically different feature for weighing and ranking pages. Link analysis works somewhat like bibliographic citation practices, such as those used by Science Citation Index. Link analysis is based on how wellconnected each page is, as defined by Hubs and Authorities, where Hub documents link to large numbers of other pages (out-links), and Authority documents are those referred to by many other pages, or have a high number of "in-links".

• **Link Popularity:** Link popularity is a major parameter Google is using. There is simple logic behind it. If other websites are linking to your site then there has to be more relevancy in your website. Popularity utilizes data on the frequency with which a page is chosen by all users as a means of predicting relevance. While popularity is a good indicator at times, it assumes that the underlying information need remains the same.

There are few more factors such as:

• Date of article published: The more recent the content the importance is more! Search engines always believe that if the content is recent then it will be more valuable for visitor than others. The engines therefore present results beginning with the most recent to the less current.

• Length: While length per se does not necessarily predict relevance, it is a factor when used to compute the relative merit of similar pages. So, in a choice between two documents both containing the same query terms, the document that contains a proportionately higher occurrence of the term relative to the length of the document is assumed more likely to be relevant.

• Proximity of query terms: When the terms in a query occur near to each other within a document; it is more likely that the document is relevant to the query than if the terms occur at greater distance. While some search engines do not recognize phrases per se in queries, some search engines clearly rank documents in results higher if the query terms occur adjacent to one another or in closer proximity, as compared to documents in which the terms occur at a distance.

• Proper nouns sometimes have higher weights, since so many searches are performed on people, places, or things. While this may be useful, if the search engine assumes that you are searching for a name instead of the same word as a normal everyday term, then the search results may be noticeably slanted.

Search Engine Spam

Search engine spamming is the unethical practice for optimizing the site to rank it high on SERP. Spamming is used to trick search engines for higher rankings with the use of some tactics such as repetitive keywords, hidden text and links etc. All search engines penalize websites that use spam. Since time immemorial --or at least since the Internet first began-- webmasters have been using these stratagems to dupe search engines into giving irrelevant pages high search engine placement.

Each search engine's objective is to produce the most relevant results to its visitors. Producing the most relevant results for any particular search query is the determining factor of being a popular search engine. Every search engine measures relevancy according to its own algorithm, thereby producing a different set of results. Search engine spam occurs if anybody tries to artificially influence a search engine's basis of calculating relevancy. Each of the major search engines provide specific guidelines describing what webmasters should and should not do to their web pages in order to achieve a better search engine ranking, though that has not always been the case.

There are overall sixteen tactics that are considered search engine spam. These techniques are:

- Keywords unrelated to site
- Redirects
- Keyword stuffing
- Mirror/duplicate content
- Tiny Text
- Doorway pages
- Link Farms
- Cloaking
- Keyword stacking
- Gibberish
- Hidden text
- Domain Spam
- Hidden links
- Mini/micro-sites
- Page Swapping (bait &switch)
- Typo spam and cyber squatting

Not to be confused with the canned, processed meat, spam is the use of redundant or unethical techniques to improve search engine placement. Fortunately, or unfortunately -- depending on your point of view-- search engines are quickly catching on. Some won't index pages believed to contain spam; others will still index, but will rank the pages lower, while others still will ban a site altogether. Of course, not all search engines take a hardline on spam. Tricks that are perfectly acceptable on one search engine may be considered spam by another.

Spamming Techniques

Invisible Text: Hiding keywords by using the same color font and background is one of the oldest tricks in the spammers' book. These days, it's also one of the most easily detected by search engines

Keyword Stuffing: Repeating keywords over and over again, usually at the bottom of the page (tailing) in tiny font or within Meta tags or other hidden tags.

Unrelated Keywords: Never use popular keywords that do not apply to your site's content. You might be able to trick a few people searching for such words into clicking at your link, but they will quickly leave your site when they see you have no info on the topic they were originally searching for. If you have a site about Medical Science and your keywords include "Shahrukh Khan" and "Britney Spears", that would be considered unrelated keywords.

Hidden Tags: The use of keywords in hidden HTML tags like comment tags, style tags, httpequiv tags, hidden value tags, alt tags, font tags, author tags, option tags, no-frames tags (on sites not using frames).

Duplicate Sites: Content duplication is considered to be search engine spamming also. Sometimes what people do is, they copy the content and name the site differently. But search engines can find it easily and they mark it as a spam. Don't duplicate a web page or doorway page, give them different names, and submit them all. Mirror pages are regarded as spam by all search engines and directories.

Link Farms: Link farm is a network of pages on one or more Web sites heavily cross-linked with each other, with the sole intention of improving the search engine ranking of those pages and sites.

Many search engines consider the use of link farms or reciprocal link generators as spam. Several search engines are known to kick out sites that participate in any link exchange program that artificially boosts link popularity. Links can be used to deliver both types of search engine spam, i.e. both content spam and Meta spam.

Link content spam: When a link exists on a page A to page B only to affect the hub component of page A or the authority component of page B, that is an example of content spam on page A. Page B is not spamming at all. Page A should receive a spam penalty. Without further evidence, page B should not receive a penalty.

Link Meta spam: When the anchor text or title text of a link either mis-describes the link target, or describes the link target using incoherent language, that is an example of link Meta spam.

Repetitive Submitting: Each search engine has its own limits on how many pages can be submitted and how often. Do not submit the same page more than once a month to the same search engine and don't submit too many pages each day. Never submit doorways to directories.

Redirects: Do not list sites using URL redirects. These include welcome.to, i.am, go.to, and others. The complete site should be hosted on the same domain as the entry page. An exception may be made for sites that include a remotely hosted chat or message board as long as the bulk of the site is hosted on its own domain. Actually redirecting of page was not developed for spam, but it is becoming popular technique for spamming. There are many means of redirecting from one Web page to another. Examples of redirection methods are HTTP 300 series redirect response codes, HTTP 400 series error vectors, META REFRESH tags and JavaScript redirects. As studied earlier these are used to move visitor from one page to another without giving them a single second. In this case the page made for search engine is a spam. Everything on it is an example of either content spam or Meta spam.

Alt Text Spamming: Tiny text consists of placing keywords and phrases in the tiniest text imaginable all over your site. Most people can't see them, but spiders can. Alt text spamming is stuffing the alt text tags (for images) with unrelated keywords or phrases.

Doorway Pages: Doorways are pages optimized only for search engine spiders in order to attract more spiders, thus more users. Usually optimized for just one word or phrase and only meant for spiders, not users.

Content Spam: It is possible when different URLs delivers same content i.e. content duplication and same URL can deliver different content as well. Both HTML and HTTP supports it and hence spamming is possible. For example, IMG support and ALT text within HTML means that image-enabled visitors to a URL will see different content to those visitors that, for various reasons, cannot view images. Whether the ability to deliver spam results in the delivery of spam is largely a matter of knowledge and ethics.

Agent based Spam: Agent based delivery is certainly not spam. But it is spam when the use of agent-based delivery to identify search engine robots by user agent and deliver unique

content to those robots. Since the content is only created for search engines and it is not visible for users, it is always spam.

IP Spam: Identification of search engine robots by IP name or address and delivery of unique content to those robots is considered to be spamming. As in agent based spam, though this technique is also spam when you deliver unique content only to search engines and not the users or visitors.

No Content: If sites do not contain any unique and relevant content to offer visitors, search engines can consider this spam. On that note, illegal content, duplicate content and sites consisting of large affiliate links are also considered to be of low value to search engine relevancy.

Meta Spam: Meta data is data that describes a resource. Meta spam is data that mis-describes a resource or describes a resource incoherently in order to manipulate a search engine's relevancy calculations.

Think again about the ALT tag. Not only does it provide content for a HTML resource, it also provides a description of an image resource. In this description capacity, to mis-describe an image or to describe it incoherently is Meta-spam. Perhaps the best examples of Meta spam at present can be found in the <head> section of HTML pages. Remember though, it's only spam if it is done purely for search engine relevancy gain.

Meta spam is more abstract than content spam. Rather than discuss it in abstract terms, we will take some examples from HTML and XML/RDF in order to illustrate Meta spam and where it differs from and crosses with content spam. Generally, anything within the section of an HTML document, or anything within the section that describes another resource, can be subverted to deliver Meta spam.
To make sure you are not spamming, you need to check a few things.

First and foremost, you should know whether your content is really valuable for your customers and visitors or not. Try and make websites according to user's tests and preferences. Always remember that, Internet users are information seekers and they want the

latest content all the time. Think and build a site as if there are no search engines and avoid automated pages. Google and many other search engines do not index auto generated pages.

Intranets

Increasingly, businesses are relying on intranets to deliver tools such as collaboration, scheduling, customer relationship management tools, and project management to increase the productivity of the organization. An **intranet** is a private network accessible only to an organization's staff. Unlike the Internet, an internal intranet provides a wide range of information and services to employees of an organization but these tools and information are unavailable to the public. A company-wide intranet is an important focal point of internal communication and collaboration, and can provide a business with a single starting point to access both internal and external resources. Larger businesses allow users within their intranet to access the public Internet through firewall servers. Because businesses have the ability to screen both incoming and outgoing traffic, they are able to keep the security of the intranet intact. In its simplest form, an intranet is established with the technologies for local area networks (LANs) and wide area networks (WANs).

1. An intranet is a corporate LAN or wide area network (WAN) that uses Internet technology and is secured behind company's firewalls (see security and protection).

2. The intranet links various servers, clients, databases, and application programs like Enterprise Resource Planning (ERP). Although intranets are developed on the same TCP/IP protocol as the Internet, they operate as a private network with limited access.

3. Only authorized employees are able to use it. Intranets are limited to information pertinent to the company and contain exclusive and often proprietary and sensitive information.

4. The firewalls protect the intranets from unauthorized outside access; the intranet can be used to enhance the communications and collaboration among authorized employees, customers, suppliers, and other business partners.

5. Since the intranet allows access through the Internet, it does not require any additional implementation of leased networks. This open and flexible connectivity is a major capability

and advantage of intranet. Intranets provide the infrastructure for many intrabusiness commerce applications.

Some of the advantages and benefits a company can realize from establishing a robust intranet are as follows.

- **Workforce productivity**. Intranets can help users to locate and view information faster and use applications relevant to their roles and responsibilities.
- **Enhanced collaboration**. Information is easily accessible by all authorized users, which enables teamwork. Being able to communicate in real-time through integrated third party tools promotes the sharing of ideas and helps boost a business' productivity
- **Time Savings**. Intranets allow organizations to distribute information to employees on an *as-needed* basis in real time. Employees may link directly to relevant information as soon as the organization makes it available on the intranet.
- **Reduced Costs**. Users can view information and data via web-browser rather than maintaining physical documents such as procedure manuals, internal phone list and requisition forms. This can potentially save the business money on printing, duplicating documents, and the environment as well as document maintenance overhead.
- **Improved Communication**. Intranets can serve as powerful tools for communication within an organization. A great real-world example of where an intranet helped a company communicate is when Nestle had a number of food processing plants in Scandinavia. Their central support system had to deal with a large number of requests for information every day. When Nestle decided to invest in an intranet, they quickly realized the savings. In fact, the savings from the reduction in calls was substantially greater than the investment in the intranet.

Extranets

In some cases organizations make the decision to allow external parties such as customers and suppliers to have access to their intranet. When these outside parties are provided access to a subset of the information accessible from an organization's intranet the intranet becomes an **extranet**. For example a large construction company may share drawings with architects or inspectors, photographs to their customers and loan documents to their bankers by implementing online applications that allow these external parties to access and even mark-up

and make changes to documents. In essence, the company will use an extranet to manage project-related communications. One of the biggest advantages of establishing an extranet is that a business can share large quantities of data using EDI or electronic data interchange. Data such as invoice and order that were traditionally transmitted via paper can now instantly be shared among organizations. Some of the most sophisticated extranets are run by large retailers like Walmart and Target who constantly transmit data via their extranet to vendors and suppliers, ensuring that merchandise arrives when it is needed, where it is needed.

1. An extranet, or "extended intranet", uses the TCP/IP protocol network of the Internet, to link intranets in different locations.

2. Extranet transmission is usually conducted over the Internet, which offers little privacy or transmission security.

3. Therefore, when using an extranet, it is necessary to improve the security of connecting portions of he Internet. This can be done by creating tunnels (see paragraph on security and protection) of secured data flows, using cryptography and authorization algorithm.

4. The Internet with tunneling technology is known as a virtually private network (VPN).

5. Extranets provide secured connectivity between corporation's intranets and the intranets of its business partners, material suppliers, financial services, government, and customers.

6. Access to intranets is usually limited by agreements of the collaborating parties, is strictly controlled, and is only available to authorized personnel.

7. The protected environment of the extranet allows groups to collaborate, sharing information exclusively, and exchanging it securely.

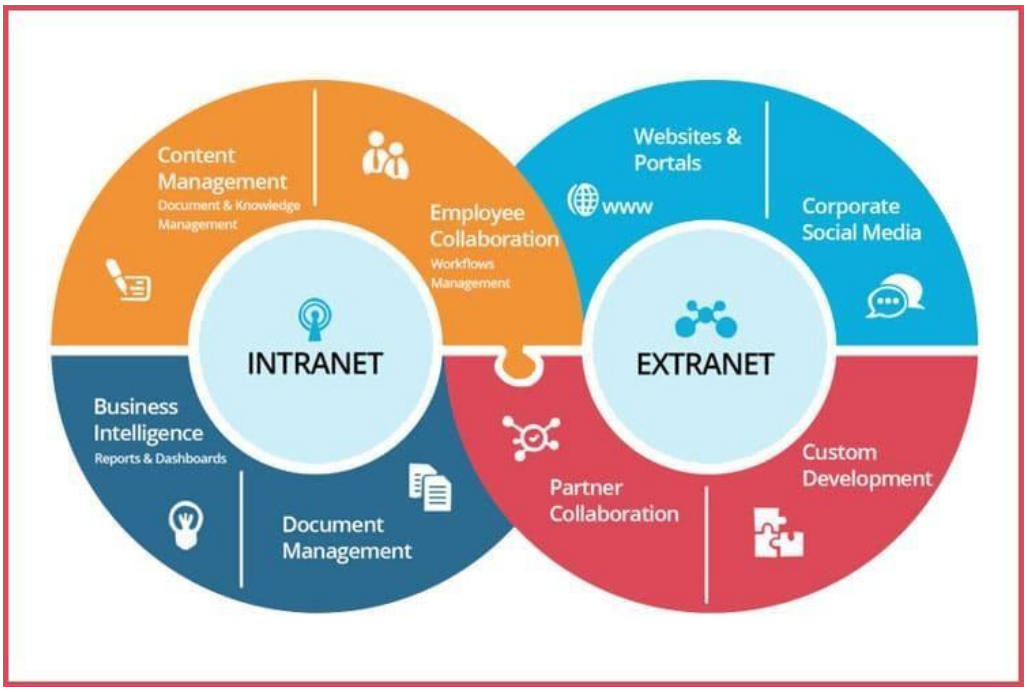
8. Since an extranet allows connectivity between businesses through the Internet, it is an open and flexible platform suitable for supply chain management.

9. To increase security, many companies replicate the database they are willing to share with their business partners and separate them physically from their regular intranets.

Like intranets, extranets have some distinct advantages for the organizations establishing them. Several of these benefits are explained below.

- **Build customer relationships.** Customers who are provided access to timely information about product availability, specifications and cost increase their efficiency. In business-to-business relationships, the more timely and accurate information a business makes available to their customers, the more likely they are to retain that business. Collaborate with other companies on joint development efforts
- **Reduced margin of error.** An extranet can reduce a company's margin of error thereby reducing or eliminating costly errors, especially with something as complex as processing orders from distributors and suppliers. Customers can be given access to their accounts to verify order history, account balances and payments.
- **Timely and accurate information.** On an extranet a business can instantly change, edit, and update sensitive information such as price lists or inventory information. Compared to typical paper-based publishing processes, an extranet offers a unique opportunity to quickly get information into the right hands before it's out-of-date.
- **Reduced inventory.** One of the greatest advantages of a business-to-business extranet is its impact on supply-chain management. By linking the inventory system directly to a supplier, businesses can process orders as soon as the system knows they are needed, thus reducing the stock a business keeps on hand and generally making the procurement process more efficient.
- **Flexibly.** A well designed extranet allows remote and mobile staff to access core business information 24 hours a day, irrespective of location. This allows employees to work remotely or respond to critical requests for information after normal working hours. As businesses expand globally, the ability to work across time zones is enhanced by the establishment of an extranet.

Whether a company is managing an intranet or extranet, both systems can raise security issues. With increased access comes an increased opportunity for security breaches. In particular, the security of extranets can be a concern when hosting valuable or proprietary information. Unless sufficient security precautions are taken, data can be breached and altered, without either the sender or the receiver being aware of the interception. The growth in the complexity of networks has increased the possible points of attack, both from within organizations and from outside the company. Fortunately, the means of protecting against hackers have also expanded in line with the technology.



With the growing dependence on the internet, most businesses use both intranet and extranet for work and communication purposes.

One can hear the terms intranet and extranet almost daily while working and communicating with team members, customers, business associates, etc. it is easy to get confused with the terms intranet and extranet.

They both are similar as their chief purpose is to improve and increase communication and collaboration. Though their role or purpose is the same on the whole intranet and extranet are very different.

Difference between Intranet and Extranet:

Before learning about the difference between intranet and extranet, let us first understand the application of intranet and extranet, their examples, and their uses.

What is Intranet?

The Intranet is a private network used by a company to enable secure communication and collaboration amongst its employees; it is also useful for storing internal information.

The prefix infra means within or inside, so in simple words, an intranet is the company's digital workspace which is centralized. It helps in streamlining every individual, document, tools, projects, etc. within the company.

What is an extranet?

Similar to the intranet, the extranet is also a private network within an organization; however, it uses the internet to connect to outsiders in a controlled manner.

Extranet helps in connecting organizations with their customers and suppliers, which helps in working collaboratively.

In simple words, it is the organization's intranet that is extended to authorized users outside the organization. A controlled, private network that enables third-party partners to access certain information without granting complete access to the organization's entire network.

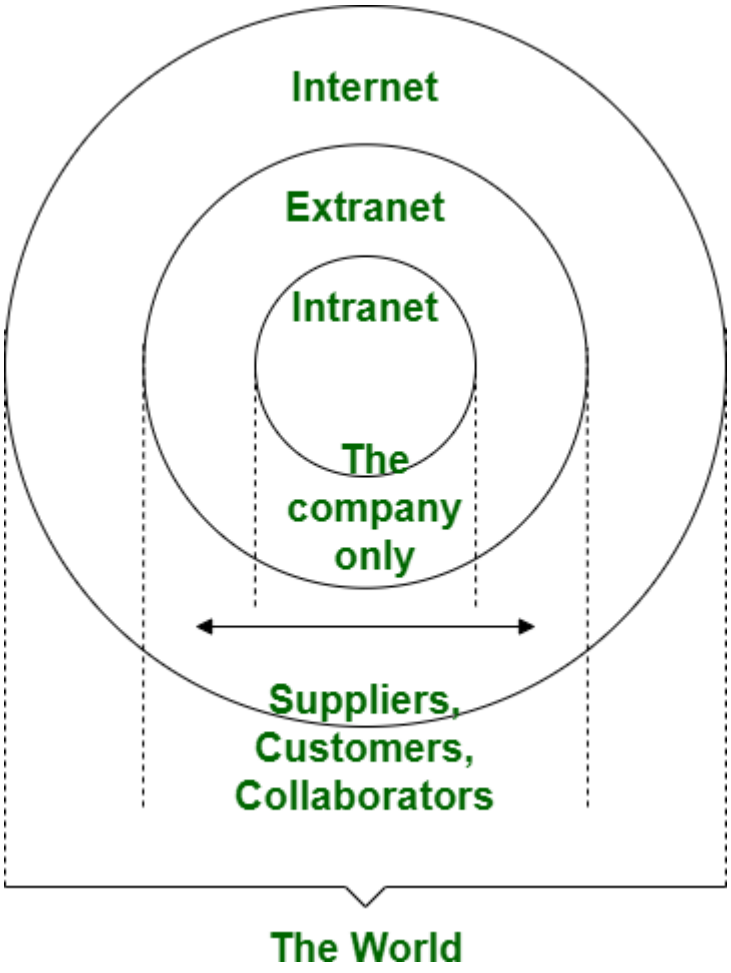
Similar to the intranet, extranet allows private communication, file, and information sharing, However, in this case, it is between authorized external partners and the organization.

Intranet vs. Extranet: what do they do?

Intranets usually begin with publishing web pages informing about the upcoming company events, company health, safety or security policies, staff newsletters, etc.

Intranet gives access to forms for claiming expenses or for requesting leaves and holidays. Having all this online on the organization's internal network helps in eliminating unwanted paperwork and speed up the process.

Additional features that are essential for the operation of the organization can be added to the intranet. It can be made into a portal that provides easy access to all the things required by the employees.



Firewalls are used for protecting the intranet from the global internet, employees can log in to the intranet only with the help of a secure password.

Employees or contractors working for the organization but are located at different locations can access the intranet by using a VPN (a virtual private network). All the communications made between the intranet and the individual's personal computer is encrypted.

Extranet provides access to a business network to people working for different organizations. For example, a business can provide access to its supplier to make online ordering, order tracking, and inventory management simpler.

Instead of sending information to the suppliers on each occasion, the extranet allows the suppliers to automatically access the required information. Another example of the extranet is the hospital appointment booking system.

Under the booking system, the doctors can access the hospital's system to make appointments for their patients. Since the communications made of the extranet can be encrypted over a VPN, it is more secure as compared to sending data over the public internet.

Intranet vs. Extranet: The key differences

- **Purpose:**

Intranet – The main purpose of the Intranet is to enable the sharing of sensitive and confidential information within the organization.

Extranet – Extranet allows communication of information between the employees of the organization and the external business associates.

- **Regulation:**

Intranet – It is regulated by the specific organization and its organizational policies.

Extranet – It is regulated by 2 or more organizations which are sharing the information and data. It is regulated by the contractual agreements made between the organizations.

- **Accessibility:**

Intranet – The content is accessible only by the members or employees of the organization.

Extranet – The content is accessible by the employees of the organization and external members who have been given access to the network.

- **Security:**

Intranet – The network is secured by a firewall.

Extranet – The network is secured through a firewall that separates the internet and extranet.

Intranet vs. Extranet: The Final Words

Intranet and Extranet both have many similarities, including their main purpose to increase communication and collaboration. Though similar they have some differences, their chief difference can be determined from the first few letters of their name.

The prefix intra means internal or inside. Which means it is an internal network dedicated to the employees of the organization.

The prefix extra means external or it refers to any contact or activity outside your business like clients, vendors, and suppliers.

SEARCH ENGINE COMPONENTS

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.

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.

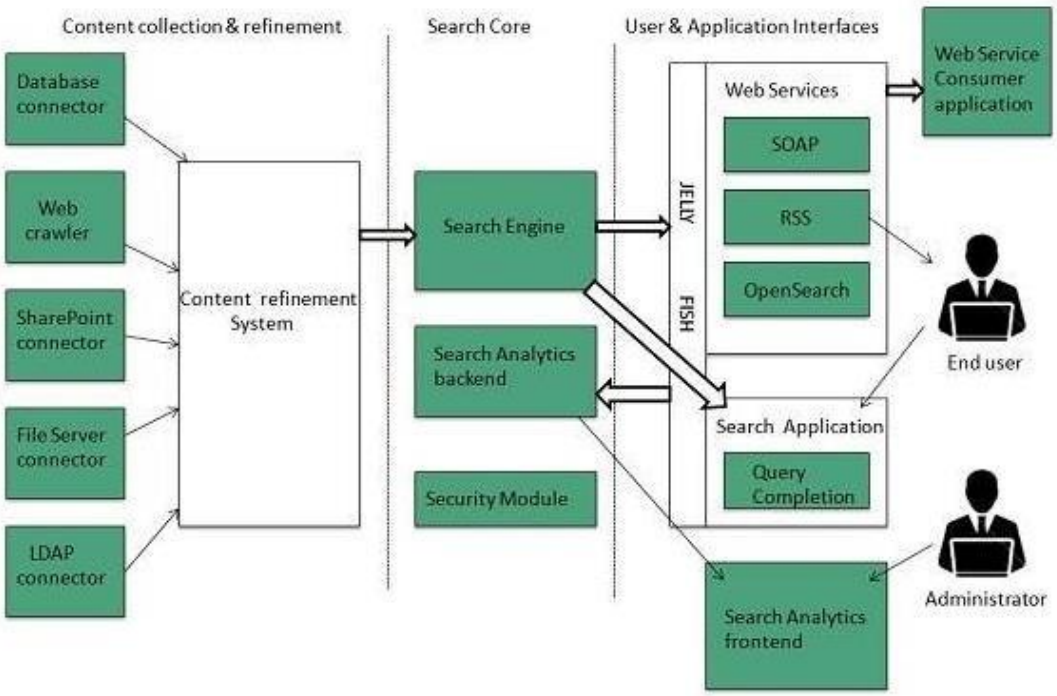
These search criteria may vary from one search engine to the other. The retrieved information is ranked according to various factors such as frequency of keywords, relevancy of information, links etc.

- User can click on any of the search results to open it.

Architecture

The search engine architecture comprises of the three basic layers listed below:

- Content collection and refinement.
- Search core
- User and application interfaces



Search Engine Processing

Indexing Process

Indexing process comprises of the following three tasks:

- Text acquisition
- Text transformation
- Index creation

Text acquisition

It identifies and stores documents for indexing.

Text Transformation

It transforms document into index terms or features.

Index Creation

It takes index terms created by text transformations and create data structures to suport fast searching.

Query Process

Query process comprises of the following three tasks:

- User interaction
- Ranking
- Evaluation

User interaction

It supporst creation and refinement of user query and displays the results.

Ranking

It uses query and indexes to create ranked list of documents.

Evaluation

It monitors and measures the effectiveness and efficiency. It is done offline.

Examples

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.
AOL.Search	It is powered by Google.
LYCOS	It is top 5 internet portal and 13th largest online property according to Media Matrix.

INTERNET SEARCH ENGINE PREREQUISITES AND SERVICES

Loosely organizing the 'net

The vast amount of information available on the Internet can be dizzying. Some authorities estimate the number of documents on the Internet to be in the range of 800 million. Others say the number is unknowable. Fortunately, there are tools available that will sort through the mass of information: search engines or search directories.

Search engines collect information from Web sites and then, more or less, just dump that information into a database. There's more information to choose from in a search engine, but it's more difficult to retrieve relevant information.

Search directories try to impose some sense of order on the information they collect and you're more likely to find information relevant to your research topic, but they don't offer the massive amounts of information that you would find with a search engine. The sites collected are viewed by humans who make decisions about what subject categories the sites might fit into.

Search engines

Search engines are really just massive databases in which information from Internet documents are stored. The information in these databases is collected using a computer

program (called a "spider" or a "robot") that scans the Internet and gathers information about individual documents. These special programs work automatically to find documents or they are asked by a creator of a Web site to visit the site to be included in a database.

When you do a search in a search engine, the order in which the results are listed also varies between search engines. Many search engines list the results using relevance ranking. Factors such as:

- how often your search terms are on the Web page;
- where they are located on the page; and,
- how many other Web pages link to the page

...influence how high on the list of hits a page is listed. Many search engines allow Web sites to pay to have their pages listed higher in the results.

There are hundreds of these search engines available on the Web, but they all work in unique ways to collect and organize the information found. The information from Web sites might be gathered from all the words in a site, just the first few sentences in the body of a site, or only from the title or metatags (hidden descriptors of a site's content). Different search engines collect different information, that's why you'll get different results from the same search from different search engines.

Search directories

Directories are best used when you are looking for information that is easily classified, such as "Universities and Colleges in Georgia." You can find the information you need without even typing in a search, but by browsing the directory, starting with a very broad subject category (Education) and working your way through the directory until you come to individual listings for schools in Georgia. You can do the usual search as well, but directories don't collect the same range of sites that a search engine would so you wouldn't be tapping into the wealth of information that you can get from a search engine.

GALILEO also has a database of useful Web sites that are evaluated by educators. These sites are not submitted by the developer nor are they harvested by spiders. They are chosen deliberately for their usefulness for research in the curriculum of the University System of Georgia.

Metasearch engines

These type of search services offer sort of a "one-stop shopping" to the Internet. You can form one search and a metasearch service will send the search to several other search engines and directories simultaneously so that you get the results from all of them in one place. The only problem with this is that you only get the first few results from each listing. If the site you're looking for happens to be listed in the 10th position in a search services results list and the metasearch engine only provides the first 5 results from that list, then you won't find the site you need. If you're only trying to get a general idea of what information is available on the Web, then a metasearch engine would be a good place to start.

Search Engine Optimization Strategies

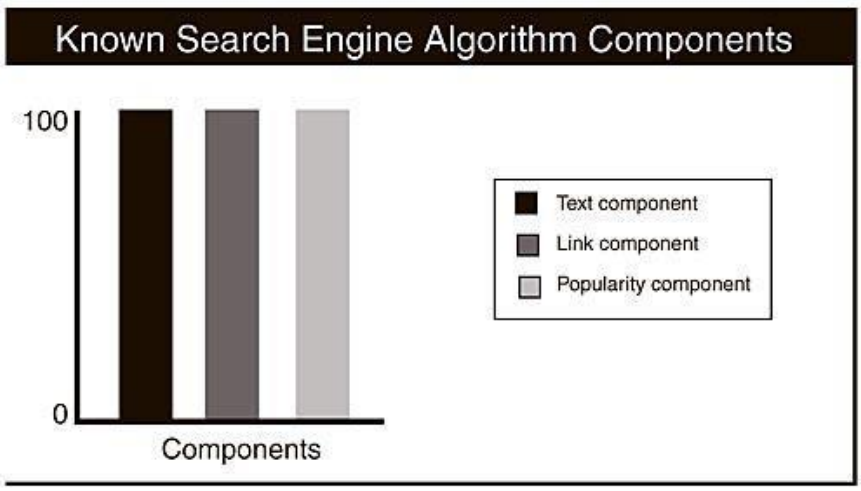
Search engine optimization is the process of designing, writing, coding (in HTML), programming, and scripting your entire web site so that there is a good chance that your web pages will appear at the top of search engine queries for your selected keywords. Optimization is a means of helping your potential customers find your web site.

To get the best overall, long-term search engine visibility, the following components must be present on a web page:

- Text component
- Link component
- Popularity component

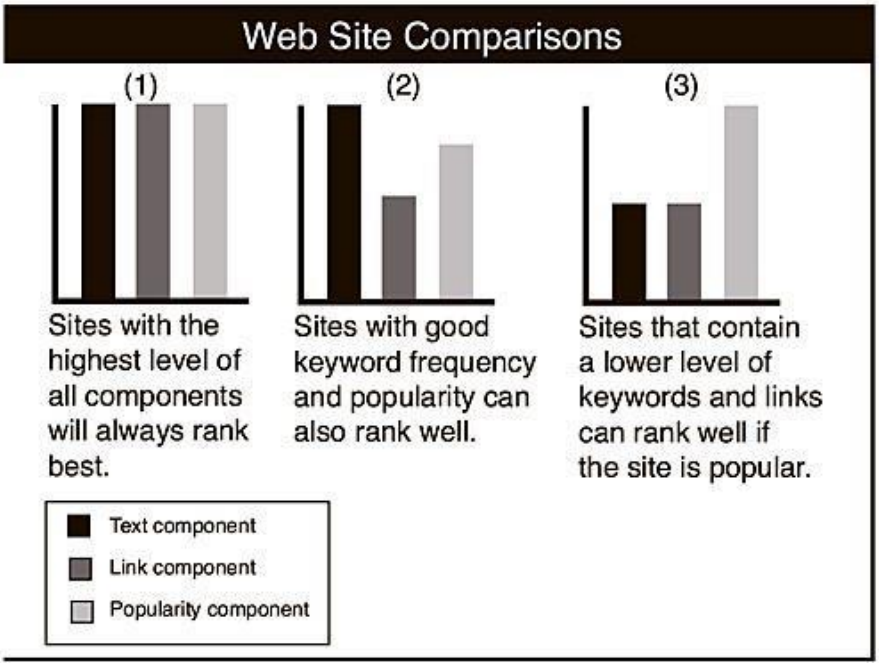
All the major search engines (Google, FAST Search, MSN Search, and other Inktomi-based engines) use these components as part of their search engine algorithm.

Known search engine algorithm components: text, link, and popularity.



Very few web pages can attain the "ideal" match for all search engine algorithms. In reality, most web pages have different combinations of these components.

Web site comparisons



Sites perform well in the search engines overall when they have (a) all the components on their web pages and (b) optimal levels of all the components.

Text Component—An Overview

Because the search engines build lists of words and phrases on URLs, it naturally follows that to do well on the search engines, you must place these words on your web pages in the strategic HTML tags.

The most important part of the text component of a search engine algorithm is keyword selection. For your target audience to find your site on the search engines, your pages must contain keyword phrases that match the phrases your target audience is typing into search queries.

After you have determined the best keyword phrases to use on your web pages, you will need to place them within your HTML tags. Different search engines do not place emphasis on the same HTML tags. For example, Inktomi places some emphasis on meta tags; Google ignores meta tags. Thus, to do well on all the search engines, it is best to place keywords in all the HTML tags possible, without keyword stuffing. Then, no matter what the search engine algorithm is, you know that your keywords are optimally placed.

Keywords need to be placed in the following places:

- Title tags
- Visible body text
- Meta tags
- Graphic images (the alternative text)

The title tag and the visible body text are the two most important places to insert keywords because all the search engines index and place significant "weight" on this text.

Keywords in Your Domain Name

Many search engine marketers believe that placing keywords in your domain name and your filenames affect search engine positioning. Some search engine marketers believe that this strategy gives a significant boost whereas others believe that the boost is miniscule.

One reason people believe the position boost is significant is that the words or phrases matching the words you typed in a query are highlighted when you view the search results. This occurrence is called *search-term highlighting* or *term highlighting*.

Search engines and directories might use term highlighting for usability purposes. The process is done dynamically using a highlighting application. This application simply takes

your query words and highlights them in the search results for quick reference. Term highlighting merely indicates that query terms were passed through the application. In other words, in search results, just because a word is highlighted in your domain name does not necessarily mean that the domain name received significant boost in search results.

Many other factors determine whether a site will rank, and the three components (text, link, and popularity) have more impact on search engine visibility than using a keyword in a domain name.

Link Component—An Overview

The strategy of placing keyword-rich text in your web pages is useless if the search engine spiders have no way of finding that text. Therefore, the way your pages are linked to each other, and the way your web site is linked to other web sites, does impact your search engine visibility.

Even though search engine spiders are powerful data-gathering programs, HTML coding or scripting can prevent a spider from crawling your pages. Examples of site navigation schemes that can be problematic include the following:

- **Poor HTML coding on all navigation schemes:** Browsers (Netscape and Explorer) can display web pages with sloppy HTML coding; search engine spiders are not as forgiving as browsers.
- **Image maps:** Many search engines do not follow the links inside image maps.
- **Frames:** Google, Inktomi, and Lycos follow links on a framed site, but the manner in which pages display in search results are not ideal.
- **JavaScript:** The major search engines do not follow many of the links, including mouseovers/rollovers, arrays, and navigation menus, embedded inside JavaScript.
- **Dynamic or database-driven web pages:** Pages that are generated through scripts or databases, or that have a ?, &, $, =, +, or % in the URL, pose problems for search engine spiders. URLs with CGI-BIN in them can also be problematic.
- **Flash:** Currently, only Google and FAST Search can follow the links embedded in Flash documents. The others cannot.

Therefore, when designing web pages, be sure to include a navigation scheme so that the spiders have the means to record the words on your web pages. Usually that means having two forms of navigation on a web site: one that pleases your target audience visually and one that the search engines spiders can follow.

Popularity Component—An Overview

The popularity component of a search engine algorithm consists of two subcomponents:

- Link popularity
- Click-through or click popularity

Attaining an optimal popularity component is not as simple as obtaining as many links as possible to a web site. The quality of the sites linking to your site holds more weight than the quantity of sites linking to your site. Because Yahoo! is one of the most frequently visited sites on the web, a link from Yahoo! to your web site carries far more weight than a link from a smaller, less visited site.

To develop effective link popularity to a site, the site should be listed in the most frequently visited directories. Yahoo!, LookSmart, and the Open Directory are examples of the most frequently visited directories.

More importantly, it can boost your search engine position if a directory that is associated with a search engine lists your site. For example, a site that is listed in LookSmart can be given higher visibility in an MSN Search.

Obtaining links from other sites is not enough to maintain optimal popularity. The major search engines and directories are measuring how often end users are clicking the links to your site and how long they are staying on your site and reading your web pages. They are also measuring how often end users return to your site. All these measurements constitute a site's click-through popularity.

The search engines and directories measure both link popularity (quality and quantity of links) and click-through popularity to determine the overall popularity component of a web site.

If a single page (web page 1) ranks well in the search engines and end users click the links to that web page and browse your site, web page 1's popularity level increases. If a different web page (web page 2) ranks well in the search engines for a different keyword phrase, web page 2's popularity level increases. The total page popularity of your site will increase your overall site's online visibility.

One of the reasons that a site's home page is more important than any other web page is that search engines assign a higher "weight" to it. In all likelihood, the home page is going to be the URL listed in the major directories, and the home page has more links to it from within the web site.

DESIGN AND EVALUATION OF SEARCH ENGINES

As the Internet provides unlimited amounts of information that can be accessed with low effort and cost, search engine represents powerful tool that assists the Internet users in their interaction with the online environment. A search engine is an information retrieval system with a set of programs with search tools used to perform searches, designed, developed, and used for finding information from the web using different strategies. Search engines perform the basic retrieval task including the acceptance of a query, a comparison with a database and the production of retrieved digital information such as text, audio, video, data and simulations. A search engine tool is a utility available in the Internet in which the user inputs specific name, subject, and/or key words for the purpose of retrieving a list of web links that match the user's query terms. The search engine facilitates the users to apply their criteria to a database to build a set of matches. The examples include Google, Yahoo!, AltaVista, Webcrawler, Lycos, Excite, Infoseek, Excite, HotBot, Bing, Ask, AOL, and other portal sites. Usually the search engine indexes with automatic software and the catalog is built manually with human input. Search engines gather web pages that form the universe from which the users retrieve information by issuing queries and the required information is retrieved by using information retrieval algorithms. Usually the search engines provide search services on the web based on directory services and query based. The directory services (e.g. yahoo) provide a hierarchical organization of resources developed by human cataloguers; the query-based services (e.g., Excite) provide broad coverage of the Internet through intensive automation of query retrieval process. Users usually search for information trying to maximize the accuracy of search outcome with minimum effort exerted to acquire it.

In particular, the following criteria are used for evaluating search engines: • value of the search results - major features (options that make the site unique, informative, and any value addition to the search results) - quality of retrieved items (proving current and authoritative information) (Jones & Timm, 2008) • convenience of various search tools • ability of search engines to locate information on the Internet • the usability of search tools (interface design -availability of basic and advanced search features with instructions for effective searches - overall ease of use) (Su, 2003b; Vaughan, 1999) - most frequently applied measure • effectiveness (number and precision or relevance of returned results) (Ziff-Davis, 1995) • search engines' capacity to retrieve the information that matches user's informational needs (Pan et al., 2007) • comprehensiveness of the Web engines by number of documents indexed (Venditto, 1996) • search capability options, how the results were displayed (readability), update frequency of information (Courtois et al., 1995) • browsability (ease of understanding results) -navigation (easy-to-use format for finding and viewing the information) (Jones & Timm, 2008) • customizability (ability to construct a search to weed out irrelevant results) • name (the name of the site), on-screen help, speed (response time/ timeliness), database coverage, number of links, accessibility and others.

Small Questions

S. No	Questions	LOCF Mapping
1.	What is a search engine?	K1
2.	Differentiate between a search engine and a directory.	K2
3.	Name the three basic components of a search engine.	K1
4.	What is search engine spam?	K1
5.	Differentiate between intranet and extranet.	K2

Big Questions

S. No	Questions	LOCF Mapping
1.	Discuss the major search engines and their market share.	K2
2.	Explain the components and working of a search engine.	K2, K3
3.	Describe the on-page and off-page factors that influence search engine ranking.	K2, K3

4.	What is search engine spam? Discuss various spamming techniques and how search engines combat them.	K2, K3
5.	Discuss the features, benefits, and applications of intranets and extranets in organizations.	K2, K3

UNIT V: Web Security

Content: Internet Security – Network Security Goals; Security Engineering; Security Requirement Engineering; Viruses, Worms, Malware, Spyware, Adware, Trojans, Botnets; Defending Against Threats; Cyber Crime; Information Technology Act 2000 and Amendment Act 2008; Firewalls, Antivirus, Anti-spyware.

UNIT 5

INTERNET AND WEB SECURITY:

Internet security is a branch of computer security. It encompasses the Internet, browser security, web site security and network security as it applies to other applications or operating systems as a whole. Its objective is to establish rules and measures to use against attacks over the Internet. The Internet is an inherently insecure channel for information exchange, with high risk of intrusion or fraud, such as phishing, online viruses, trojans, ransomware and worms.

We entrust networks with our social life (e.g. social networks), our money (e.g. Internet banking), our health (e.g. electronic health records) and more. The high value combined with the fact that networks inherently connect multiple parties, each with their own interests, means that they are a natural target for attackers. Ensuring security of networks is thus of

paramount importance. Yet what does that exactly mean? Which security goals of which stakeholders are important to achieve?

On a computer network with di_erent interconnected systems security is not only important, but also hard to achieve. Not only do we need to consider threats on our own (local) system but also on all systems connected to it, as well as the connections themselves. Where we may have some trust in our own system we likely will not trust all systems on the network and their users. The interests of the other parties on the network may be completely different than ours. As we have seen in our security analysis, convicting interests lead to (security) risks. In this chapter we will look at some of specific computer network (protocol) related threats and corresponding countermeasures.

In our discussion we try to answer the following questions:

- Why are networks and web applications so vulnerable?
- How to achieve network and web security goals?
- How to approach di_erent attacker models?

We start by looking at threats at the lower network layers, and move up the protocol stack illustrating that risks exist at each layer and also in moving between the layers. Next we touch on (distributed) denial of service attacks that are focused on breaking the security attribute `availability'. We then move to the top of the network protocol stack and we treat the application layer separately, looking in particular at web services and related vulnerabilities. Having described the threats we discuss some key network security technologies aiming to address these threats such as security protocols and intrusion detection and prevention systems (_rewalls, intrusion detection systems, virus detection, etc.).

NETWORK SECURITY

NETWORK SECURITY GOALS

Networked systems (simple apps, complex networks, complete IT infrastructures) operate in environments involving different interconnected parties each with their own agenda (goals), which may not match with the goals of other parties of the system as whole. As such, it is essential to also consider the security requirements of systems (i.e. what should not go wrong), not only their functional requirements (i.e. what the systems should achieve).

"Is your system secure?" What does this question actually mean; does it mean that nobody but you can use it; can throw it out the window; can keep you from using it...? Security requirements are expressed in terms of security attributes that express goals that one may want to achieve to call a system 'secure'. The most commonly used and widely accepted security attributes are Confidentiality, i.e. 'my information stays secret', Integrity, i.e. 'my information stays correct', and Availability, i.e. 'I can get at my information' (sometimes called the C-I-A triad). Of course these concepts can also refer to resources or system aspects other than just 'information'.

In addition to Confidentiality, Integrity and Availability ('C-I-A') other security attributes are sometimes formulated. Closely related but not usually called a security attribute is Privacy, i.e. 'information about me is not misused'. Note the difference between Confidentiality and

Privacy: where confidentiality requires data that you possess to remain secret, privacy deals with data about you that may be in the hands of others. While 'who gets the data' is a key question in confidentiality, the purpose for which data is used is a key ingredient for privacy.

Other examples of security attributes are: Authenticity, i.e. 'is this information authentic (i.e. of undisputed origin)', Non-repudiation, i.e. 'is this information undeniable' and Accountability, i.e. 'is the information provider accountable (can we punish the provider if the information is incorrect)'. Authenticity is different from integrity in that it focuses on data coming from the 'correct' source rather than on data not being changed along the way. A signature on a contract would be an example of a way to achieve non-repudiation; you cannot later deny agreeing to the conditions in the contract. The relation between accountability and non-repudiation is similar to that between integrity and authenticity; non repudiation can be an important part of achieving accountability but is by itself not sufficient.

The security requirements together with the security policies of a system tell you what attributes should be achieved when (in which context). The requirements will typically say what security attributes should be achieved by which components and/or for what type of resources (e.g. confidential database entries should only be readable by user with the right clearance). Security policies detail this e.g. by stating what type of data is confidential and what (types of) users have clearance. Security requirements are an integral part of the design of the system while changes of policies is typically taken into account and should not invalidate the design. Note, however, that the term security policy is widely used and the exact interpretation varies. It could be a high level textual description meant to be understood and applied by human beings, e.g. "all personal identifiable information must only be read when needed to provide a service" to low level computer readable information e.g. "drwxr-xr-x"5 . Translating high level policies into a systems design along with low level policies is an important step of creating a secure system.

The exact meaning of a security policy can be given within a security model; a (formal) framework to express and interpret policies. For example, the Unix file permission given above can be interpreted as a relation between Users, Groups, Objects and Permissions: An object (e.g. a directory) has an owner user and a group (an additional part of the security policy) and the owner of the object has read, write and execute permission, while members of the group as well as other users have only read and execute permission.

Attacker models can be general, e.g. IBM's classification of attackers into three categories (clever outsiders, knowledgeable insiders and funded organizations); or formal, e.g. those used in analysis of cryptographic algorithms (e.g. Chosen-Plaintext-Attack (CPA) where the attacker is able to get encryptions of plain text she has chosen). Any security analysis will need both the security goals (attributes/policy) and the attacker model. Sometimes these are left implicit but they remain key ingredients; the question 'is this system secure?' has no meaning without them. Not properly considering them is a common cause of security problems.

SECURITY ENGINEERING

A chain is no stronger than its weakest link. This is also the case for the security of a system. Consider for example the following aspects of a system and some potential issues.

Design

There is no hope of having a secure system if the system design does not address security goals or worse has inherent features/goals that imply security problems. As an example consider the Windows Meta File (WMF) where arbitrary code execution, a clear security risk, is a design feature.

As another example one can consider the Internet; initially the Internet linked a group of trusted systems. Security goals that are very important now were thus not under consideration in its design, e.g. no protection of content, any computer can claim to have an IP, no authentication of DNS, etc. Of course there are currently security mechanisms (IPsec, HTTPS, etc.) that try to remedy this but 'add on security' is always problematic - security needs to be considered from the start.

Software quality

A perfect design does not help if the implementation is flawed. Often security issues are caused by software bugs with buffer overflow vulnerabilities being one of the major issues. In buffer overflow attacks input from an untrusted source is written into a buffer without the bounds of the buffer being checked. This causes the untrusted data to be written to places it is not supposed to go; it may overwrite a return address on the stack, causing a jump to an attacker selected location at the end of the current routine.

The problem of software bugs is not solved easily; e.g. an unsolved buffer overflow vulnerability was reported in Windows 7 and in January 2011 Microsoft shipped fixes for 22 vulnerabilities. The 'heartbleed bug' is a recent example of a software flaw related security incident with wide media coverage. Note that software and systems evolve. It is not the case that each round of patching brings us closer to a final secure and bug free system.

Security Tool Selection

Choose your crypto well, especially if you are a mafia boss. From a news article: "...He apparently wrote notes to his henchmen using a modified form of the Caesar Cipher, which was easily cracked by the police and resulted in further arrests of collaborators..." Clearly here the selected security tool was grossly insufficient to reach the security goal.

This is an extreme example but often inappropriate security tools are used or tools are used well past their 'best before/replace by' date such as the hash function MD5, which has been known to be vulnerable for a long time but is only slowly being phased out. Using 'home-made' crypto solutions instead of tried and proven standard algorithms would also fit in this category. A good practice is to leave design of crypto to the experts; obscurity of a design is not a good replacement for their experience and expertise.

System usage

Even a perfectly designed and implemented security architecture (should one ever be created) is of no help if it is not used correctly. USB data sticks that offer encryption of their content are readily available and company policy may state that the encryption of such sticks should be used. However, if the user does not enable this feature this is all for nothing.

Users have different priorities; e.g. ease of use; and many do not use security features or will even try to work around them if they interfere with what they are trying to do.

Of course these are only examples and there are many more aspects of a system where a weak link in the security chain may occur. The key points are that one needs to consider the system as a whole and consider security from the start.

We have already seen some security tools (means) above and later we will try to add key tools to this toolbox, focusing on network scenarios. Cryptography is an important part of this toolbox. However recall that security tools by themselves do not make the system secure. A common claim 'the data is secure because it is encrypted' is by itself meaningless and may

even indicate that the security goals and the attacker model have not been sufficiently considered. For instance, encryption offers no protection against inside attackers who have access to the key. A good security design determines what security tools need to be employed where and when, considering the security requirements and the effects (including trade-offs) different tools have on these requirements.

SECURITY REQUIREMENT ENGINEERING

As already mentioned several times (and as will be repeated later in more detail), to really evaluate the security of a system you have to consider it as a whole, know the security goals and the potential threats against these goals. To gather these we need to perform Security Requirement Engineering. Throughout the design, implementation, deployment and use of a system we should consider the requirements that the users will have from the system and how attackers will try to exploit the system. Based on this we can come up with and/or evaluate a security design, which combines several security solutions to achieve the best possible trade-offs.

Here we shortly cover one example security requirement approach. Other approaches may work just as well, what is important is that the security requirements are considered throughout in a structured and consistent way.

Identify actors and goals

The first step in gathering the requirements is determining the stakeholders and their interests. The stakeholders are those parties with a legitimate interest in the system that we are designing. Their interest and goals thus have to be considered (though not necessarily completely reached - we may need to make trade-offs between the different goals of the participants).

The stakeholders and their interests become the initial actors and goals in the requirements gathering process. If an agent has the right capabilities, it may adopt a goal, i.e. take responsibility to achieve it. If an agent does not adopt the goal it may be delegated to other agents (either existing or new) or be split into new sub goals. Agents do not work in isolation; agents and their goals may depend on/interact with each other. These dependencies should be

identified and could lead to new goals and/or agents. They also lead to potential vulnerabilities, e.g. when agents' goals conflict.

So far the process matches a typical functional requirement engineering approach. In order to deal with security requirements we also need to consider attackers and possible attacks on the system.

Identify attackers, vulnerabilities and attacks

Outsiders may try to attack our system and they need to be considered along with their goals. However, also the risk of attacks by insiders needs to be accounted for. Each agent in the system could potentially become an attacker, using its capabilities and place in the system to reach their goals at the expense of the goals of other agents. Both types of attackers are modeled as agents in the system but with malicious intent as their goal.

Based on vulnerabilities and the malicious intent of attacker agents we identify potential attacks and assign countermeasures to protect against such attacks. The countermeasures themselves may lead to new actors/goals and/or open the possibility for new attacks which need to be considered. Refinement of the system continues until all goals have been assigned, dependencies taken into account, and vulnerabilities addressed.

VIRUSES

A computer virus is a malicious piece of executable code that propagates typically by attaching itself to a host document that will generally be an executable file.

In the context of talking about viruses, the word "host" means a document or a file. As you'll recall from our earlier discussions, in the context of computer networking protocols, a "host" is typically a digital device capable of communicating with other devices. Even more specifically, in the context of networking protocols, a host is whatever is identified by a network address, like the IP address.

Typical hosts for computer viruses are:

- Boot sectors on disks and other storage media
- Executable files for system administration (such as the batch files in Windows machines, shell script files in Unix, etc.)
- – Documents that are allowed to contain macros (such as PDF files, Microsoft Word documents, Excel spreadsheets, Access database files, etc.)

Any operating system that allows third-party programs to run can support viruses. Because of the way permissions work in Unix/Linux systems, it is more difficult for a virus to wreak havoc in such machines. Let's say that a virus embedded itself into one of your script files. The virus code will execute only with the permissions that are assigned to you. For example, if you do not have the permission to read or modify a certain system file, the virus code will be constrained by the same restriction.

At the least, a virus will duplicate itself when it attaches itself to another host document, that is, to another executable file. But the important thing to note is that this copy does not have to be an exact replica of itself. In order to make more difficult its detection by pattern matching, a virus may alter itself when it propagates from host to host. In most cases, the changes made to the virus code are simple, such as rearrangement of the order independent instructions, etc. Viruses that are capable of changing themselves are called mutating viruses.

Computer viruses need to know if a potential host is already infected, since otherwise the size of an infected file could grow without bounds through repeated infection. Viruses typically place a signature (such as a string that is an impossible date) at a specific location in the file for this purpose.

Most commonly, the execution of a particular instance of a virus (in a specific host file) will come to an end when the host file has finished execution. However, it is possible for a more vicious virus to create a continuously running program in the background. To escape detection, the more sophisticated viruses encrypt themselves with keys that change with each infection. What stays constant in such viruses is the decryption routine. The payload part of a virus is that portion of the code that is not related to propagation or concealment.

WORMS

The main difference between a virus and a worm is that a worm does not need a host document. In other words, a worm does not need to attach itself to another program. In that sense, a worm is self-contained. On its own, a worm is able to send copies of itself to other machines over a network. Therefore, whereas a worm can harm a network and consume network bandwidth, the damage caused by a virus is mostly local to a machine.

But note that a lot of people use the terms ‘virus’ and ‘worm’ synonymously. That is particularly the case with the vendors of anti-virus software. A commercial anti-virus program is supposed to catch both viruses and worms. Since, by definition, a worm is supposed to hop from machine to machine on its own, it needs to come equipped with considerable networking support. With regard to autonomous network hopping, the important question to raise is: What does it mean for a program to hop from machine to machine?

A program may hop from one machine to another by a variety of means that include:
By using the remote shell facilities, as provided by, say, ssh, rsh, rexec, etc., in Unix, to execute a command on the remote machine. If the target machine can be compromised in this manner, the intruder could install a small bootstrap program on the target machine that could bring in the rest of the malicious software. By cracking the passwords and logging in as a regular user on a remote machine. Password crackers can take advantage of the people’s tendency to keep their passwords as simple as possible (under the prevailing policies concerning the length and complexity of the words).

By using buffer overflow vulnerabilities in networking software. In networking with sockets, a client socket initiates a communication link with a server by sending a request to a server socket that is constantly listening for such requests. If the server socket code is vulnerable to buffer overflow or other stack corruption possibilities, an attacker could manipulate that into the execution of certain system functions on the server machine that would allow the attacker’s code to be downloaded into the server machine.

In all cases, the extent of harm that a worm can carry out would depend on the privileges accorded to the guise under which the worm programs are executing. So if a worm manages to guess someone’s password on a remote machine (and that someone does not have elevated privileges), the extent of any harm done might be minimal. Nevertheless, even when no local “harm” is done, a propagating worm can bog down a network and, if the propagation is fast enough, can cause a shutdown of the machines on the network. This can happen particularly when the worm is not smart enough to keep a machine from getting reinfected repeatedly and simultaneously. Machines can only support a certain

maximum number of processes running simultaneously. Thus, even “harmless” worms can cause a lot of harm by bringing a network down to its knees.

Malware

Malware, short for “malicious software,” is a blanket term that refers to a wide variety of software programs designed to do damage or do other unwanted actions to a computer, server or computer network common examples include viruses, spyware and trojan horses. Malware can slow down or crash your device or delete files. Criminals often use malware to send spam, obtain personal and financial information and even steal your identity. Above we have looked at some specific vulnerability of networks and the machines on a network. Different types of malware exploit such weaknesses to infiltrate the system, replicate, spread and achieve some malicious goal.

Trojans are legitimate looking programs (or other content) that actually carry malicious code inside. Viruses can replicate, usually involving some action like running an infected program (like biological viruses need a host organism, computer viruses infect programs). Worms are able to replicate by themselves, without the need of human action. Well known worms, such as conficker, spread around networks (e.g. the Internet) exploiting vulnerabilities of network services and machines. Adware is a type of software that automatically shows advertisements or redirects your searches to advertising websites, in order to generate revenues for its creator. (This is not necessarily malicious, however many types of adware try to hide on your system, avoid uninstall and/or gather information without your consent.) Classification of malware into these categories, however, is sometimes difficult and terms are commonly mixed, using e.g. virus for any type of malware.

While some hacks, viruses and worms may have been `just to see what I can do', i.e. idealistic, to demonstrate the vulnerability or simple vandalism, modern malware is for the most part targeting big businesses or are even used for digital warfare. The conficker worm, for instance, installs scareware (showing pop-ups to get user to buy a fake anti-virus solution) and creates a botnet; a network of machines under control of the attacker, which can then be

used for sending spam, distributed DoS attacks, renting out to others, etc. The worm has an update mechanism used to download its software for its malicious activities as well as updates to its spreading mechanism and protection against updates of anti-virus software that would be able to protect against it. (Zero-day) vulnerabilities, exploits, virus building kits and botnets are commodities that are traded on the black market.

One thus does not even have to create one's own malware or botnet; it is possible to buy or rent infected machines. (See for example Metasploit www.metasploit.com for hacking made easy.) Botnets are controlled through command and control centers. By using a C&C center to drive a botnet, the malware can easily be updated and adapted, making it hard to take down the network of bots. To prevent the C&C center itself from being taken down, it is located in countries where there are no laws to easily do this, its location is hidden e.g. within a list of addresses, using anonymous services in TOR3 and/or redundancy is used so a new C&C can easily be created at a di_erent location. As in many security areas there is an arms race between taking out botnet C&C centers and new infections, botnets and control methods appearing.

Of course, with all the value a botnet represents there are those that will try to take it over, either to dismantle/study it (e.g. [24]) or to use it for their own illegitimate agenda. It has also been suggested to take this defense strategy a step further; use weaknesses in infected machine to force installation of patches, removing of and protecting against malware but there are many moral, legal, practical and technical issues with such an approach. Related to this is use of `hacking' by authorities which is also the subject of debate what actions are justi_ed, should e.g. `hacking' by the police be allowed (and if so to what degree, in which circumstances)? For example German federal court rejects hacking by police in general though leaves open some possibilities to do so for example with respect to threats that endanger public safety such as terrorism.

SPYWARE

Spyware is a type of malware that attaches itself and hides on a computer’s operating system without your permission to make unwanted changes to your user experience. It can be

used to spy on your online activity and may generate unwanted advertisements or make your browser display certain website sites or search results.

DEFENDING AGAINST NETWORK SECURITY THREATS

Above we have already mentioned several methods of countering specific network threats. Here we treat three general categories of network defense technologies: Virus scanners, Network Intrusion Detection Systems (IDS, NIDS), and firewalls. Virus scanners try to detect and disable or remove (clean) malware, intrusion detection systems try to detect malicious traffic and firewalls try to block malicious traffic.

A key problem that we see returning in each of these technologies, as well as many other security mechanisms, (input filtering, access control, biometrics, etc.) is the need to distinguish `good' cases (code, traffic, input, requests, measurements, etc.) from `bad' cases. In general we have two ways of approaching this:
Whitelisting; by using an access control matrix to describe what traffic (or . . .) is `good'. Anything else is considered `bad'.
Blacklisting; by describing what traffic is `bad'. Any other traffic is considered `good'. Many virus scanners work in this way; they look for (patterns of) known viruses in programs.
The AC matrix tries to simply list all allowed (`good') cases. As we have seen this quickly becomes unmanageable, thus prompting the introduction of more advanced specification methods such as RBAC and XACML. In other settings there are simply way too many possibilities to create a simple list. We thus need to express a model for `good' or `bad' in a more ecient way. For example input filtering against SQL is a form of blacklisting where we use a set of rules, for example, \the ' symbol is not allowed in a user input used to construct an SQL query" or replace the offending character with a safe encoding.)

CYBER CRIME AND INFORMATION TECHNOLOGY ACT

The Term ‘Cyber Crime’ needs no introduction in today’s E-world. In this world, where everything is available at a click, crimes are also been committed at a click. Cyber Crime thus is the darker side of technology. It is a Crime where the computer is either a tool or a target.

The term WWW which stands for World Wide Web has now become World Wide Worry because of mushroom growth in cyber-crimes.

Crime in a developing nation is a hindrance to its development. It not only adversely affects all the members of the society but it also pulls down the economic growth of the country. Computer Technology provided a boost to the human life. It made the life of human being easier and comfortable. It not only added speed to the life of human being, but it also added accuracy and efficiency. But this computer was exploited by the criminals. This illegal use of computers for commission of crime leads to Cyber Crime. To combat Cyber Crime India got armed herself with The Information Technology Act 2000. This act got drastically amended in year 2008. The Amended Information Technology Act is not only effective than the previous Act it is more powerful and stringent than the previous one. II IMPACT OF CO

IMPACT OF COMPUTER ON HUMAN LIFE

Change is the rule of universe. Nothing in this world is static and technology is providing a pace to this change. The highlight of this era is e-governance. That means the government is available to its citizens at just a click. A farmer is not required to the village officer for obtaining his property extract, its available online to him. The long queues for paying bills are becoming history, people are preferring to pay bills online. Ecommerce is becoming a part of business. Shopping on internet through e-commerce website is becoming a trend. Telegram technology has already said good bye to the World, because mobile is available in every pocket.

The impact of globalization and computerisation is phenomenal. It is an era when now we can dream of a paperless world. The United Nations Commission on International Trade Law (UNCITRAL) adopted the Model Law on e-commerce in 1996. The General Assembly of United Nations passed a resolution in January 1997 inter alia, recommending all States in the UN to give favourable considerations to the said Model Law, which provides for recognition to electronic records and according it the same treatment like a paper communication and record.

CYBER CRIME

Cyber Crime is the darker side of technology. The term ‘Cyber Crime’ finds no mention either in The Information Technology Act 2000 or in any legislation of the Country. Cyber

Crime is not different than the traditional crime. The only difference is that in Cyber Crime the computer technology is involved. This can be explained by following instance;

Traditional Theft : A thief enters in B's house and steals an object kept in the house.

Hacking : A Cyber Criminal sitting in his own house, through his computer hacks the computer of B and steals the data saved in B's computer without physically touching the computer or entering in B's house

Hence Cyber Crime is a Computer related crime. The I.T. Act, 2000 defines the terms access in computer network in section 2(a), computer in section 2(i), computer network in section (2j), data in section 2(0) and information in section 2(v). These are all the necessary ingredients that are useful to technically understand the concept of Cyber Crime. In a cyber crime, computer or the data are the target or the object of offence or a tool in committing some other offence. The definition of term computer elaborates that computer is not only the computer or laptop on our tables, as per the definition computer means any electronic, magnetic, optical or other high speed data processing devise of system which performs logical, arithmetic and memory function by manipulations of electronic, magnetic or optical impulses, and includes all input, output, processing, storage, computer software or communication facilities which are connected or related to the computer in a computer system or computer network. Thus the definition is much wider to include mobile phones, automatic washing machines, micro wave ovens etc...

PREAMBLE OF INFORMATION TECHNOLOGY ACT

The Preamble of the I. T. Act reflects the objectives with which The Government of India enacted The Act. The objectives of the Act are;

1. To provide legal recognition for transactions carried out by means of electronic data interchange and other means of electronic communication, commonly referred to as "electronic commerce", which involve the use of alternatives to paper-based methods of communication and storage of information,

2. To facilitate electronic filing of documents with the Government agencies and further to amend the Indian Penal Code, the Indian Evidence Act, 1872, the Bankers' Books Evidence

Act, 1891 and the Reserve Bank of India Act, 1934 and for matters connected therewith or incidental thereto."

The Information Technology Act, 2000, was thus passed as the Act No.21 of 2000. The I. T. Act got President assent on 9th June 2000 and it was made effective from 17th October 2000. By adopting this Cyber Legislation India became the 12th Nation in the world to adopt a Cyber Law regime during 2000.

SILENT FEATURES OF INFORMATION TECHNOLOGY ACT

The silent features of the Act are;

- The Act gives legal recognition of Electronic Documents.
- The Act gives legal recognition of Digital Signatures.
- It describes and elaborates Offenses, penalties and Contraventions.
- It gives outline of the Justice Dispensation Systems for cyber crimes.
- The Act also provides for the constitution of the Cyber Regulations Advisory Committee, which shall advice the government as regards any rules, or for any other purpose connected with the said act.
- The said Act also proposed to amend to; The Indian Penal Code, 1860, The Indian Evidence Act, 1872, The Bankers' Books Evidence Act, 1891, The Reserve Bank of India Act, 1934 etc...

SCHEME OF THE INFORMATION TECHNOLOGY ACT

The I. T. Act is spread in total 13 chapters. There are total 90 sections, the last four sections namely sections 91 to 94 in the I. T. Act 2000 dealt with the amendments to the Indian Penal Code 1860, The Indian Evidence Act 1872, The Bankers' Books Evidence Act 1891 and the Reserve Bank of India Act 1934 were deleted. The I. T. Act commences with Preliminary aspect in Chapter 1, which deals with the short, title, extent, commencement and application of the Act in Section 1. Section 2 provides Definition. Chapter 2 deals with authentication of electronic records, digital signatures, electronic signatures etc. Thereafter elaborate procedures for certifying authorities (for digital certificates as per IT Act -2000 and since replaced by electronic signatures in the ITAA -2008) are been provided. Chapter XI deals

with offences and penalties. A series of offences have been provided along with punishment in this part of The Act. Thereafter the provisions about due diligence, role of intermediaries and some miscellaneous provisions are been stated.

The Act is embedded with two schedules. The First Schedule deals with Documents or Transactions to which the Act shall not apply. The Second Schedule deals with electronic signature or electronic authentication technique and procedure. The Third and Fourth Schedule are omitted.

APPLICATION OF THE INFORMATION TECHNOLOGY ACT

As per Section 1 of The I. T. Act, the Act extends to the whole of India and except as otherwise provided, it applies to also any offence or contravention there under committed outside India by any person. As per sub clause (4) of Section 1, Nothing in this Act shall apply to documents or transactions specified in First Schedule. As per this first schedule following are the documents or transactions to which the Act shall not Apply;

1. Negotiable Instrument (Other than a cheque) as defined in section 13 of the Negotiable Instruments Act, 1881;

2. A power-of-attorney as defined in section 1A of the Powers-of-Attorney Act, 1882;

3. A trust as defined in section 3 of the Indian Trusts Act, 1882;

4. A will as defined in clause (h) of section 2 of the Indian Succession Act, 1925 including any other testamentary disposition;

5. Any contract for the sale or conveyance of immovable property or any interest in such property;

6. Any such class of documents or transactions as may be notified by the Central Government.

THE I. T. ACT, 2000 PROVIDES FOR FOLLOWING DEFINITION;

TABLE I

Section 2(a)	Access		**Section 2(r)**	Electronic Form
Section 2(b)	Addressee		**Section 2(s)**	Electronic Gazette
Section 2(c)	Adjudicating Officer		**Section 2(t)**	Electronic Record
Section 2(d)	Affixing		**Section 2(ta)**	Electronic Signature
Section 2(e)	Appropriate Authority		**Section 2(tb)**	Electronic Signature Certificate
Section 2(f)	Asymmetric Crypto System		**Section 2(u)**	Function
Section 2(g)	Certifying Authority		**Section 2(ua)**	Indian Computer Emergency Response Team
Section 2(h)	Certification Practice Statement		**Section 2(v)**	Information
Section 2(ha)	Communication Device		**Section 2(w)**	Intermediary
Section 2(i)	Computer		**Section 2(x)**	Key Pair
Section 2(j)	Computer Network		**Section 2(y)**	Law
Section 2(k)	Computer Resource		**Section 2(z)**	Licence
Section 2(l)	Computer System		**Section 2(za)**	Originator
Section 2(m)	Controller		**Section 2(zb)**	Prescribed
Section 2(n)	Cyber Appellate Tribunal		**Section 2(zc)**	Private key
Section 2(na)	Cyber Café		**Section 2(zd)**	Public key
Section 2(nb)	Cyber Security		**Section 2(ze)**	Secure system
Section 2(o)	Data		**Section 2(zf)**	Security procedure
Section 2(p)	Digital Signature		**Section 2(zg)**	Subscriber
Section 2(q)	Digital Signature Certificate		**Section 2(zh)**	Verify

AMENDMENT BROUGHT IN THE I. T. ACT BY AMENDMENT ACT OF 2008

Cyber Crime is a technology related offence. Technology is never static. It keeps on changing and getting better and better. At the same time Cyber Criminals are exploiting this advanced technology to discover sophisticated ways of committing crime. The Information Technology Act is the saviour in the nation to combat cyber crimes. Thus as the criminals are keeping pace with the advancement in technology, it is equally important for the Law to keep itself update with the recent trends in commission of crime and advancement in technology. With the same intention the Amendment Act of 2008 brought sweeping changes in the old I. T. Act. To overcome the lacuna of old I. T. Act, many bodies, teams of technical experts and advisory groups were construed. They studied the cyber legislations in other foreign countries and recent trend in cyber crime scenario. Their recommendations were scrutinized and the Parliament of India came up with Information Technology Amendment Act 2008. It was placed in the Parliament and passed without much debate. This Amendment Act got the nod of President 05 th February 2009. The Amendments were made effective on 27th October 2009.

HIGHLIGHTS OF THE AMENDMENT ACT, 2008

The newly amendment Act came with following highlights;

- It focuses on privacy issues.
- It focuses on Information Security.

- It came with surveillance on Cyber Cases.
- The Concept of Digital Signature was elaborated.
- It clarified reasonable security practices for corporate.
- Role of Intermediaries were focuses.
- It came with the Indian Computer Emergency Response Team.
- New faces of Cyber Crime were added.
- Powers were given to Inspector to investigate cyber crimes as against only to DSP.
- Severe Punishments and fine were added.

DIGITAL SIGNATURE TO ELECTRONIC SIGNATURE

The term 'Digital Signature' was defined in the old I. T. Act, 2000. This term was replaced by 'Electronic Signature' by the amending Act of I. T. Act, 2008. Certainly the concept of Electronic Signature is much wider than term Digital Signature. Section 3 of the Act provides for authentication of Electronic Records by affixing his Digital Signature. It shall be effected by the use of asymmetric crypto system and hash function which envelop and transform the initial electronic record into another electronic record. By the Amendment Act of 2008 Section 3(A) was embedded in the Act. The newly added provision provides for authentication of electronic record by electronic signature or electronic authentication technique which is, considered reliable and may be specified in the second schedule. Sub Clause (2) provides the circumstances in which the electronic signature or electronic authentication technique shall be considered reliable.

According to the **United Nations Commission on International Trade Law (UNCITRAL)** electronic authentication and signature methods may be classified into the following categories;

1. Those based on the knowledge of the used or the recipient i.e passwords, personal identification numbers (PINs) etc…

2. Those bases on the physical features of the user i.e. biometrics.

3. Those based on the possession of an object by the user i.e. codes or other information stored on a magnetic card.

4. Types of authentication and signature methods that, without falling under any of the above categories might also be used to indicate the originator of an electronic communication (Such as a facsimile of a handwritten signature, or a name typed at the bottom of an electronic message).

According to the **UNCITRAL MODEL LAW** on Electronic Signatures, technologies currently in use include;

1. Digital Signature within a public key infrastructure (PKI)

2. Biometric Device.

3. PINs

4. Passwords

5. Scanned handwritten signature

6. Signature by Digital Pen

7. Clickable "OK" or "I Accept" or "I Agree" click boxes

ELECTRONIC GOVERNANCE

In this era of computer where every word is getting prefixed by word 'E', Government of India is also not lacking behind and to provide its services to the citizens at their finger tips the Government is also turning in E-Governance. E-Governance is nothing but providing Government Services cheaper, faster and efficiently to the citizens through internet and computer. The Information Technology Act, 2000 gives recognition to the Electronic Governance. Chapter III, Section 4 to Section 10-A, of the Act provides for the provisions regarding Electronic Governance. Section 4 and 5 gives Legal Recognition to electronic records and electronic signatures. Section 6 of the Act authenticates use of electronic record and electronic signatures in Government and its agencies.

The aim electronic government is to ensure transparency in Government. It also makes the Government accessible to the citizen residing in the most remote village of the country.

COMPOUNDING OF OFFFENCES

As per Section 77-A of the I. T. Act, any Court of competent jurisdiction may compound offences, other than offences for which the punishment for life or imprisonment for a term exceeding three years has been provided under the Act.

No offence shall be compounded if;

1. The accused is, by reason of his previous conviction, is liable to either enhanced punishment or to the punishment of different kind; OR

2. Offence affects the socio economic conditions of the country; OR

3. Offence has been committed against a child below the age of 18 years; OR

4. Offence has been committed against a woman.

The person accused of an offence under this Act may file an application for compounding in the Court in which offence is pending for trial and the provisions of Sections 265-B and 265-C of Cr. P. C. Shall apply.

The Information Technology Act is the sole savior to combat cyber crime in nature. Though offences where computer is either tool or target also falls under the Indian Penal Code and other legislation of the Nation, but this Act is a special act to tackle the problem of Cyber Crime. The Act was sharpend by the Amendment Act of 2008, yet the Act is still in its budding stage. There is grave underreporting of cyber crimes in the nation. Cyber Crime is committed every now and then, but is hardly reported. The cases of cyber crime that reaches to the Court of Law are therefore very few. There are practical difficulties in collecting, storing and appreciating Digital Evidence. Thus the Act has miles to go and promises to keep of the victim of cyber crimes.

FIREWALLS

We already briefly mentioned firewalls when taking about IP spoofing. As we have seen above, the effect of different attacks and the ease with which they are employed depends on the capabilities of an attacker. A local attacker is able to do much more easily cause more damage then a remote attacker. As such it makes sense to try to build barriers between different parts of the network. Firewalls do exactly this

by filtering the traffic between two networks, e.g. an organization's LAN and the Internet, an organization's web services and its intranet, a single PC and the intranet, etc.

Filtering can happen at different layers of abstraction (and thus at different network layers). For example a basic packet filtering firewall (working at the network layer) can help against IP spoofing of local addresses by outside attackers and can block access to ports (and services) that should not be accessible. If working at the TCP level, it will likely need to remember which sessions are open to be able to distinguish real responses from spoofed messages; a stateful firewall.

Increased filter complexity can have a significant impact on the resources needed and thus the performance of the firewall. Going up to the application layer one can try to block dangerous data or known threats; such as remove active components from web pages, macros from word documents, block downloaded files containing viruses, tag spam and phishing emails, etc. Of course this greatly increases the complexity of the firewall; instead of looking at single packets one needs to understand the protocol being used, extract and reconstruct the data being sent, interpret and evaluate the data to determine whether it is harmful.

A main difference between firewalls and intrusion detection systems is that the former actually blocks traffic considered to be malicious. With an IDS the operator still needs to respond to deal with the attack (for example by defining a new firewall rule.) The use of firewalls thus has a direct impact on availability and usability; if we block traffic then availability and usability will go down. For example, by blocking port 22 of all machines except the public SSH server you help protect the network (e.g. a mis-configured machine vulnerable on port 22 would not be accessible from the outside) but also disallow other computers on the network from offering SSH connections. Also, the firewall will impact network performance with more advanced filtering requiring more effort thus adding to cost and/or leading to further loss of performance.

A good policy (a model of `good' and/or `bad' traffic) is thus needed to make firewalls useful; one that implements optimal trade-offs between protecting against network risks and keeping network services available. The risks, ways to detect attacks and network services needed change dynamically, making maintaining an effective firewall challenging.

The use of anomaly based approaches is uncommon when actually blocking traffic; not only is there the risk of a high false positive rate but also the reason why a particular message did not arrive might become unclear. Knowing a set of rules for the firewall a user can understand why certain connections do not work and what to change in the rules to make them work.

However, understanding a machine-learned model of used for anomaly detection, let alone updating it, is a lot more difficult. Note that the firewall can only inspect the traffic that it can see. If data is encrypted (a best practice for securing a connection) this limits the possibilities for the firewall.

Summarizing, a firewall is a very useful tool and is employed by nearly anyone operating a network. They are even built into operating systems to protect the (network consisting of only the) PC from the network it is connected to. Still, given its inherent limitations it is by itself not sufficient to protect a local network. Thus it is typically combined with IDS to find threats on the local network and anti-virus on the end-points; the machines on the network.

ANTIVIRUS

Software that is created specifically to help detect, prevent and remove malware (malicious software). Antivirus is a kind of software used to prevent, scan, detect and delete viruses from a computer. Once installed, most antivirus software runs automatically in the background to provide real-time protection against virus attacks.

Comprehensive virus protection programs help protect your files and hardware from malware such as worms, Trojan horses and spyware, and may also offer additional protection such as customizable firewalls and website blocking.

ANTI-SPYWARE

Spyware, a type of malware that allows people to spy on your activities on your devices, allows unsavory individuals to gain access to that information! Obviously, that is an outcome that few people would appreciate.

As such, a number of companies have released **anti-spyware** software that is designed to combat this threat. In this lesson, we will look at who needs anti-spyware, what it does, and different sources of reliable programs.

While many people could benefit from the advantages that anti-spyware programs have to offer, a few could certainly use them more than most. After all, there is a gulf of difference with respect to risk between the iPhone user who only uses her device to call and text and the PC user who downloads at will. After all, many mobile phones, including both the iPhone and Android phones, have sophisticated anti-spyware capabilities included in the operating system, and Android phones even let you download additional anti-spyware apps. The same can also be said for computers that only access downloads from reliable locations. If the only thing you download is music from one of the major stores, your risk of spyware is much less than someone who frequently downloads attachments from e-mails. However, in either case, anti-spyware is a relatively easy way to make your computer more secure.

Anti-spyware programs have two principal tasks. The first is to prevent spyware from becoming active on your computer in the first place. As such, it runs in the background, looking for any applications that are trying to sneak their way onto your computer or phone. It then alerts you of any points of concern.

Secondly, anti-spyware programs seek to find and delete spyware programs that have made it onto your device. To this end, they scan your hard drive and other disks periodically, making sure that nothing is hiding. This is one of the reasons that, upon initially using an anti-spyware program, many applications ask you to run the scan once, then restart the computer and run it again.

Small Questions

S. No	Questions	LOCF Mapping
1.	What are the three main security goals (CIA triad)?	K1
2.	Differentiate between a virus and a worm.	K2
3.	What is malware?	K1
4.	Define spyware.	K1
5.	What is a firewall?	K1

Big Questions

S. No	Questions	LOCF Mapping
1.	Discuss the key network security goals and their importance.	K2
2.	Explain the different types of malware (viruses, worms, trojans, spyware, adware) with their characteristics.	K2
3.	Describe the provisions of the Information Technology Act 2000 and its Amendment Act 2008.	K2
4.	Discuss the role of firewalls, antivirus, and anti-spyware in protecting networks and systems.	K2, K3
5.	Analyze the impact of cyber crime on society and the legal framework to combat it in India.	K4, K5