FUNDAMENTALS OF WEB

Dr P.V. Praveen Sundar,
Assistant Professor,
Department of Computer Science
Adhiparasakthi College of Arts & Science,
Kalavai.

22-01-2021

- I MCA

Internet

- Internet is a network of network that connects computers all over the world.

- A Computer network is a group of computers linked to each other that enables the computer to communicate with another computer and share its resources, data, and applications.

- Internet is a collection of government, academic, commercial, individual and other sites.

History of Internet

- In 1960, Advanced Research Projects Agency (ARPA) sponsored a conference at which several dozen ARPA-funded graduate students were brought together at the University of Illinois at Urbana-Champaign to meet and share ideas.

- During this conference, ARPA rolled out the blueprints for networking the main computer systems of about a dozen ARPA-funded universities and research institutions. They were to be connected with communications lines operating at a then-stunning 56 Kbps (i.e., 56,000 bits per second)—this at a time when most people (of the few who could) were connecting over telephone lines to computers at a rate of 110 bits per second.

- **Paul Baran** proposed a distributed network based on data in message blocks in the early 1960s

- There was great excitement at the conference. Researchers at Harvard talked about communicating with the UNIVAC 1108 "supercomputer" at the University of Utah to handle calculations related to their computer graphics research.

- Many other interesting possibilities were raised. Academic research about to take a giant leap forward.

- Shortly after this conference, ARPA proceeded to implement the ARPANET, which eventually evolved into today's Internet.

- **Donald Davies** conceived of packet switching in 1965 at the National Physical Laboratory (NPL) and proposed building a national commercial data network in the UK.

- In 1969, the Department of Defense (DoD) the USA created a small network of four computers called ARPANET (**A**dvanced **R**esearch **P**rojects **A**gency **N**etwork).

- This Network was set up for military purposes.

- The Primary goal of ARPANET was to allow multiple users to send and receive information simultaneously over the communication path.

☐ The network operated with a technique called Packet Switching.

☐ **Packet switching is a method of grouping data that is transmitted over a digital network into packets.**

☐ The packets contained an address, error-control, and sequencing information.

☐ The address information allowed packets to be routed to their destinations.

☐ The sequencing information helped in reassembling the packets which, because of complex routing mechanisms, could actually arrive out of order into their original order for the presentation to the recipient.

- Packets from different senders were intermixed on the same lines. This packet-switching technique greatly reduced transmission costs, as compared with the cost of dedicated communications lines.

- The network was designed to operate without centralized control.

- If a portion of the network failed, the remaining working portions would still route packets from senders to receivers over alternative paths for reliability.

- The protocol for communicating over the ARPANET became known as the Transmission Control Protocol(TCP). TCP ensured that messages were properly routed from sender to receiver and that they arrived intact.

- On October 29, 1969, ARPAnet delivered its first message: a "node-to-node" communication from one computer to another. (The first computer was located in a research lab at the University of California, Los Angeles, and the second was at Stanford University, California; each one was the size of a small house.

- The ARPANET was successful, and many universities joined the network.

- This ARPANET was divided into two parts MILNET and ARPANET.

- MILNET was used for military-related sites and ARPANET for non-military related sites.

- The technology continued to grow in the 1970s after scientists Robert Kahn and Vinton Cerf developed Transmission Control Protocol and Internet Protocol, or TCP/IP, a communications model that set standards for how data could be transmitted between multiple networks.

- During the 1970s networks like BITNET and USENET came into being.

- Around the 1980s NSFNET (National Science Foundation Network) was created.

- In the early 1980s, the National Science Foundation (NSF) funded national supercomputing centers at several universities in the United States and provided interconnectivity in 1986 with the NSFNET project. Thus creating network access to these supercomputer sites for research and academic organizations in the United States. International connections to NSFNET, the emergence of architecture such as the Domain Name System, and the adoption of TCP/IP internationally on existing networks marked the beginnings of the Internet.

- ARPANET adopted TCP/IP on January 1, 1983, and from there, researchers began to assemble the "network of networks" that became the modern Internet.

W3C

- In October 1994, Tim Berners Lee founded an organization called the World Wide Web Consortium (W3C).

- The W3C is also a standardization organization. Web technologies standardized by the W3C are called Recommendations.

- W3C Recommendations include the Extensible HyperText Markup Language (XHTML), Cascading Style Sheets (CSS), HyperText Markup Language (HTML—now considered a "legacy" technology) and the Extensible Markup Language (XML).

- A recommendation is not an actual software product, but a document that specifies a technology's role, syntax rules and so forth.

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

☐ Increasing amounts of data are transmitted at higher and higher speeds over fiber-optic networks operating at 1 Gbit/s, 10 Gbit/s, or more.

☐ The Internet's takeover of the global communication landscape was rapid in historical terms: it only communicated 1% of the information flowing through two-way telecommunications networks in the year 1993, 51% by 2000, and more than 97% of the telecommunicated information by 2007.

☐ The Internet continues to grow, driven by ever greater amounts of online information, commerce, entertainment, and social networking services. However, the future of the global network may be shaped by regional differences.

Need for World Wide Web

- The online world then took on a more recognizable form in 1990, when computer scientist Tim Berners-Lee invented the World Wide Web (linking hypertext documents into an information system, accessible from any node on the network).

- While it's often confused with the internet itself, the web is actually just the most common means of accessing data online in the form of websites and hyperlinks.

- The web helped popularize the internet among the public, and served as a crucial step in developing the vast trove of information that most of us now access on a daily basis.

- In 1989, Tim Berners-Lee of CERN (the European Organization for Nuclear Research) began to develop a technology for sharing information via hyperlinked text documents.

- Berners-Lee called his invention the HyperText Markup Language (HTML).

- He also wrote communication protocols to form the backbone of his new information system, which he called the World Wide Web.

- In particular, he wrote the Hypertext Transfer Protocol (HTTP)—a communications protocol used to send information over the web.

- Web use exploded with the availability in 1993 of the Mosaic browser, which featured a user-friendly graphical interface. Marc Andreessen, whose team at NCSA developed Mosaic, went on to found Netscape, the company that many people credit with initiating the explosive Internet economy of the late 1990s.

World Wide Web

- WWW is the part of the internet that supports multimedia and contains of a collection of linked documents.

- Web is a system of interlinked hypertext documents accessed via the internet.

- Web is an application that uses the internet for communications with TCP/IP as the underlying transport mechanism.

- A web is a huge collection of pages of information linked to each other around the globe.

□ Web uses the HTTP to transmit the data.

□ Web pages are stored on web server and send them to a client computer as and when it requests for them.

□ Internally, a web page is a computer file stored on the disk of the server. The file contains tags written in the codified form. These tags decide how the file would look when displayed on the screen. The website address is called as Uniform Resource Locator (URL).

□ A website is a collection of web pages. These pages on a website are stored digitally on a web server.
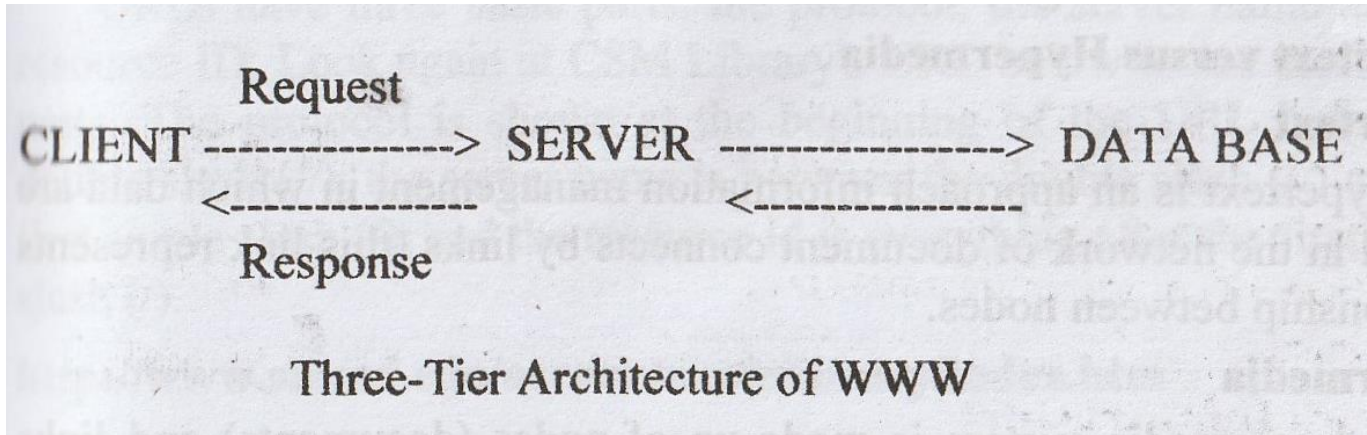
World Wide Web as Architecture

- WWW is a hypertext-based system that provides a uniform and a user-friendly interface for accessing the resources on the Internet.

- It is an information space in which the items of interest, referred to as resources, are identified by global identifiers called Uniform Resource Identifiers (URI).

- The architecture of WWW is two tiered. It consists of the client and the server. The client (web browser) requests for a web page. This page is retrieved from the server. The architecture depends on three key standards:
  - HTML for encoding document content,
  - Uniform Resource Locator (URL) for naming remote information objects in a global namespace, and
  - HTTP for staging the transfer.

□ The following figure shows the two tired architecture of WWW.

Request

CLIENT ------------------ > SERVER

<------------------

Response

Two-Tier Architecture of WWW
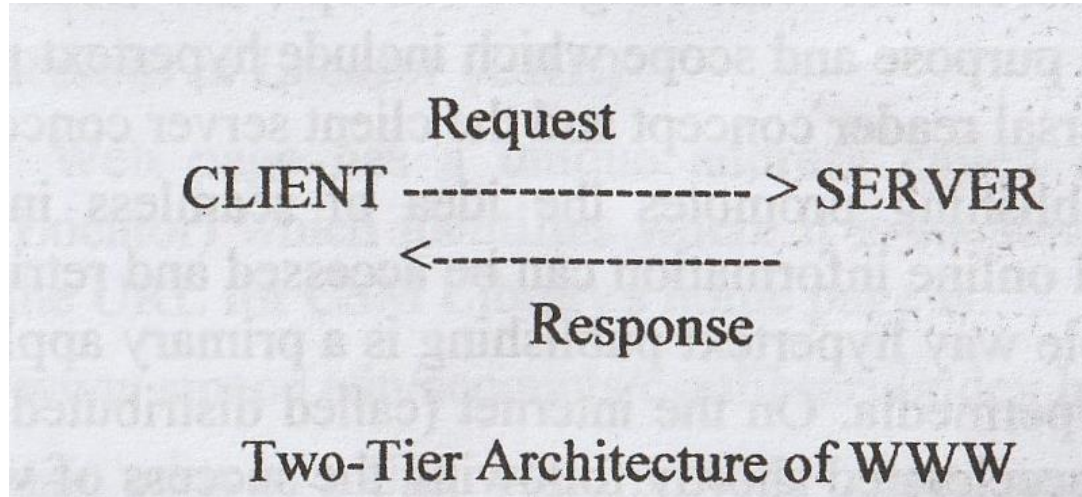
□ If the web pages are interacting with the database, then the architecture becomes three tiered, as shown in the following figure.

```
                        Request
CLIENT ----------------> SERVER ------------------> DATA BASE
       <---------------          <-----------------
        Response

         Three-Tier Architecture of WWW
```

Working of the WWW

- The WWW works on a Client-server approach. Whenever the user wants to retrieve a webpage, the www works as follows:

  - A user enters the URL of the webpage in the address bar of the web browser.

  - The web browser requests the Domain Name Server for the IP address corresponding to www.google.com.

  - After receiving the IP address, the browser sends the request for the webpage to the Internet using HTTP protocol which specific the way the browser and Web Server communicates.

- The Internet Routers send the request to the intended web server.

- Then the web server receives the request using HTTP protocol. It then examines the hard disk or memory and if the requested file is found it returns it back to the web browser and closes the Http connection.

- The Web browser then interprets the file and displays contents of the webpage in the browser window.

HTTP

- Hypertext Transfer Protocol(HTTP) is an internet standard or set of rules that allows the exchange of information on the World Wide Web.

- HTTP defines how messages are formatted and transmitted and what actions Web servers and browsers should take in response to various commands.

- For example, when anyone enter a URL, in browser, this actually sends an HTTP command to the Web server directing it to fetch and transmit the requested Web page. Sending and receiving messages can be done through HTTP.

Web 1.0

- Web 1.0 refers to the first stage of the World Wide Web evolution.

- Earlier, there were only few content creators in Web 1.0 with the huge majority of users who are consumers of content.

- Personal web pages were common, consisting mainly of static pages hosted on ISP-run web servers, or on free web hosting services.

- According to Berners-Lee, is the "Read-only Web." In other words, the early web allowed us to search for information and read it. There was very little in the way of user interaction or content generation.

- In Web 1.0 advertisements on websites while surfing the internet is banned.

- Also, in Web 1.0, Ofoto is an online digital photography website, on which user could store, share, view and print digital pictures.

- Web 1.0 is a content delivery network (CDN) which enables to showcase the piece of information on the websites. It can be used as personal websites.

- It costs to user as per pages viewed. It has directories which enable user to retrieve a particular piece of information.

Four design essentials of a Web 1.0 site include:

1. Static pages.

2. Content is served from the server's file-system.

3. Pages built using Server Side Includes or Common Gateway Interface (CGI).

4. Frames and Tables used to position and align the elements on a page.

Web 2.0

- Web 2.0 refers to World Wide Website which highlight user-generated content, usability and interoperability for end users.

- Web 2.0 is also called participative social web. It does not refer to a modification to any technical specification, but to modify in the way Web pages are designed and used.

- According to Tim Berners Lee, Web 2.0 is considered as "Read Write" Web.

- The transition is beneficial but it does not seem that when the changes are occurred. An interaction and collaboration with each other is allowed by Web 2.0 in a social media dialogue as creator of user-generated content in a virtual community.

- Web 1.0 is enhanced version of Web 2.0.

- The web browser technologies are used in Web 2.0 development and it includes AJAX and JavaScript frameworks.

- Recently, AJAX and JavaScript frameworks have become a very popular means of creating web 2.0 sites.

Five major features of Web 2.0 –

- Free sorting of information, permits users to retrieve and classify the information collectively.

- Dynamic content that is responsive to user input.

- Information flows between site owner and site users by means of evaluation & online commenting.

- Developed APIs to allow self-usage, such as by a software application.

- Web access leads to concern different, from the traditional Internet user base to a wider variety of users.

Usage of Web 2.0 –

The social Web contains a number of online tools and platforms where people share their perspectives, opinions, thoughts and experiences. Web 2.0 applications tend to interact much more with the end user. As such, the end user is not only a user of the application but also a participant by these 8 tools mentioned below:

1. Podcasting
2. Blogging
3. Tagging
4. Curating with RSS
5. Social bookmarking
6. Social networking
7. Social media
8. Web content voting

Web 3.0

- It refers the evolution of web utilization and interaction which includes altering the Web into a database. In enables the upgradation of back-end of the web, after a long time of focus on the front-end (Web 2.0 has mainly been about AJAX, tagging, and another front-end user-experience innovation). Web 3.0 is a term which is used to describe many evolutions of web usage and interaction among several paths. In this, data isn't owned but instead shared, where services show different views for the same web / the same data.

- The Semantic Web (3.0) promises to establish "the world's information" in more reasonable way than Google can ever attain with their existing engine schema. This is particularly true from the perspective of machine conception as opposed to human understanding. The Semantic Web necessitates the use of a declarative ontological language like OWL to produce domain-specific ontologies that machines can use to reason about information and make new conclusions, not simply match keywords.

Web 3.0 Features

- **Semantic Web:** The succeeding evolution of the Web involves the Semantic Web. The semantic web improves web technologies in demand to create, share and connect content through search and analysis based on the capability to comprehend the meaning of words, rather than on keywords or numbers.

- **Artificial Intelligence:** Combining this capability with natural language processing, in Web 3.0, computers can distinguish information like humans in order to provide faster and more relevant results. They become more intelligent to fulfil the requirements of users.

- **3D Graphics:** The three-dimensional design is being used widely in websites and services in Web 3.0. Museum guides, computer games, ecommerce, geospatial contexts, etc. are all examples that use 3D graphics.

□ **Connectivity:** With Web 3.0, information is more connected thanks to semantic metadata. As a result, the user experience evolves to another level of connectivity that leverages all the available information.

□ **Ubiquity:** Content is accessible by multiple applications, every device is connected to the web, the services can be used everywhere.

Web 1.0	Web 2.0	Web 3.0
Mostly Read-Only	Wildly Read-Write	Portable and Personal
Company Focus	Community Focus	Individual Focus
Home Pages	Blogs / Wikis	Live-streams / Waves
Owning Content	Sharing Content	Consolidating Content
Web Forms	Web Applications	Smart Applications
Directories	Tagging	User Behaviour
Page Views	Cost Per Click	User Engagement
Banner Advertising	Interactive Advertising	Behavioural Advertising

Web Browser

- A web browser (commonly referred to as a browser) is a software application for accessing information on the World Wide Web.

- When a user requests a web page from a particular website, the web browser retrieves the necessary content from a web server and then displays the page on the user's device.

- **Note:** A web browser is not the same thing as a search engine, though the two are often confused. For a user, a search engine is just a website that provides links to other websites. However, to connect to a website's server and display its web pages, a user must have a web browser installed.

☐ Some of the Frequently used Browsers are

- ☐ Google Chrome
- ☐ Mozilla Firefox
- ☐ Internet Explorer
- ☐ Microsoft Edge
- ☐ Opera
- ☐ Netscape navigator
- ☐ Safari

Function

- The purpose of a web browser is to fetch information resources from the Web and display them on a user's device.

- This process begins when the user inputs a Uniform Resource Locator (URL), such as https://www.apcas.in/, into the browser.

- Virtually all URLs on the Web start with either http: or https: which means the browser will retrieve them with the Hypertext Transfer Protocol (HTTP).

- In the case of https:, the communication between the browser and the web server is encrypted for the purposes of security and privacy.

□ Once a web page has been retrieved, the browser's rendering engine displays it on the user's device. This includes image and video formats supported by the browser.

□ Web pages usually contain hyperlinks to other pages and resources. Each link contains a URL, and when it is clicked or tapped, the browser navigates to the new resource. Thus the process of bringing content to the user begins again.

□ Most browsers use an internal cache of web page resources to improve loading times for subsequent visits to the same page. The cache can store many items, such as large images, so they do not need to be downloaded from the server again. Cached items are usually only stored for as long as the web server stipulates in its HTTP response messages.

Settings

□ Web browsers can typically be configured with a built-in menu. Depending on the browser, the menu may be named Settings, Options, or Preferences.

□ The menu has different types of settings. For example, users can change their home page and default search engine. They also can change default web page colors and fonts. Various network connectivity and privacy settings are also usually available.

Privacy

- During the course of browsing, browsing history cookies received from various websites are stored by the browser. Some of them contain login credentials or site preferences. However, others are used for tracking user behavior over long periods of time, so browsers typically provide settings for removing cookies when exiting the browser. Finer-grained management of cookies usually requires a browser extension.

Features

- Hyperlinks
- Tabbed Browsing
- Using the History Feature
- AutoComplete
- Off-Line Browsing
- Downloads
- Viewing Source Code
- Bookmarks

Web Server

- A Web Server is a computer that is dedicated to provide web services to clients on the internet.

- A Web server is a dedicated computer that uses HTTP (Hypertext Transfer Protocol) and other protocols to respond to client requests made over the World Wide Web.

- The main job of a web server is to display website content through storing, processing and delivering webpages to users.

- Basically **web server** is used to host the **web** sites but there exists other **web servers** also such as gaming, storage, FTP, email etc.

Web - Server Types

- **Apache HTTP Server:** This is the most popular web server in the world developed by the Apache Software Foundation.

- Apache web server is an open source software and can be installed on almost all operating systems including Linux, Unix, Windows, FreeBSD, Mac OS X and more.

- About 60% of the web server machines run the Apache Web Server.

Internet Information Services:

☐ The Internet Information Server (IIS) is a high-performance Web Server from Microsoft.

☐ This web server runs on Windows NT/2000 and 2003 platforms (and may be on upcoming new Windows version also).

☐ IIS comes bundled with Windows NT/2000 and 2003; Because IIS is tightly integrated with the operating system, so it is relatively easy to administer it.

Lighttpd:

- lighttpd: pronounced lighty is also a free web server that is distributed with the FreeBSD operating system.

- This open source web server is fast, secure and consumes much less CPU power.

- Lighttpd can also run on Windows, Mac OS X, Linux and Solaris operating systems.

Sun Java System Web Server:

- This web server from Sun Microsystems is suited for medium and large websites.

- Though the server is free it is not open source.

- However, it runs on Windows, Linux and Unix platforms.

- The Sun Java System web server supports various languages, scripts and technologies required for Web 2.0 such as JSP, Java Servlets, PHP, Perl, Python, Ruby on Rails, ASP and Coldfusion etc.

Jigsaw Server:

- ☐ Jigsaw (W3C's Server) comes from the World Wide Web Consortium.

- ☐ It is open source and free and can run on various platforms like Linux, Unix, Windows, Mac OS X Free BSD etc.

- ☐ Jigsaw has been written in Java and can run CGI scripts and PHP programs.

Uniform Resource Locator (URL)

- Every Web page has a unique address called a URL (Uniform Resource Locator) which identifies where it is located on the Web.

- For example, the URL for APCAS home page is: http://www.apcas.in/home/

- The basic parts of a URL often provide "clues" to where a web page originates and who might be responsible for the information at that page or site.

- URLs have three basic parts: the protocol, the server name and the resource ID.

- Look again at APCAS's URL below to see these three parts.

- The protocol is shown at the beginning of the URL before the double slash (//); the server name is between the double slash (//) and the first single slash (/); and the resource id is everything after the first single slash(/).

    http://www.apcas.in/home/

    protocol |Server Name |resource id

- Let's examine each part of this URL:

First part: protocol (http://)

- The protocol identifies the method (set of rules) by which the resource is transmitted. All Web pages use Hypertext Transfer Protocol (HTTP). Thus, all web URL's begin with http://.

- **Second Part: Server Name**
- The server name identifies the computer on which the resource is found.
- This part of the URL commonly identifies which company, agency or organization may be either directly responsible for the information, or is simply providing the computer space where the information is stored.
- Web server names often begin with the letters www, but not always.
- The server name always ends with a dot and a three-letter or two-letter extension called the domain name.

- The domain is important because it usually identifies the type of organization that created or sponsored the resource. Sometimes it indicates the country where the server is located.

- The most common domain names are:
  - .com which identifies company or commercial sites
  - .org for non-profit organization sites
  - .edu for educational sites
  - .gov for government sites
  - .net for Internet service providers or other types of networks.

- If the domain name is two letters, it identifies a country, e.g. .us for the United States,

  - .uk for the United Kingdom,

  - .au for Australia,

  - .mx for Mexico or

  - .ca for Canada.

- The server name for our college website is: www.apcas.in. The server name may also be the name of a website.

- Websites can be either all of the pages on one server(computer) or all of the pages under a specific sub directory on a server.

- **Third part: resource ID (home/)**
- The resource ID is the name of the file for the page and any directories or subdirectories under which it is stored on the specified computer.
- The resource ID for our College's homepage is: /home.htm. (A "homepage" is the opening or main page for any web site that provides links to all of the other pages on the site.)
- The part of the resource ID after the last slash (/) is the file name for the specific page or other resource.
- The file name ends with a three or four letter designation that specifies the file type (e.g., .htm or .html for A standard Web page, .jpg or .gif for common graphic files.)

Protocols

- ☐ Network protocols are formal standards and policies made up of rules, procedures and formats that determine how two computers communicate with one another over a network.

- ☐ Those rules indicates the formatting, ordering, and error checking data sent across a network.

- ☐ Protocol determine how the sending device indicates that it has finished sending a message and how the receiving device will indicate that it has received (or not received) the message.

Contd.,

- They incorporate all the processes requirement and constraints of initiating and accomplishing communication between computers, routers, servers and other network enabled devices.

- Network protocols must be confirmed and installed by the sender and receiver to ensure network\data communication.

- It also applies software and hardware nodes that communicate on a network.

- Some protocols also support message acknowledgement and data compression designed for reliable and high performance network communication.

Elements of Protocol

☐ The key elements of protocols are as follows:

☐ **Syntax:** Syntax concerns the format of the data blocks. It refers to the structure or format of the data that is the order in which they are represented.

☐ Example: A Simple protocol may expect the first 8 bits of data to be the address of the sender, the second 8 bits to be the address of the receiver, and the rest of the stream to be the message itself.

Contd.,

- **Semantics:** it refers to the meaning of each section of bits. How is a specified pattern to be interpreted; and what action is to be taken based on that interpretation.

- For e.g. Does an address identify the route to be taken or the final destination of the message.

- **Timing:** Timing refers to two characteristics:
  - When the data should be sent.
  - How fast they can be sent.

- For E.g.: if a sender produces data at 100 Mbps but the receiver can process data at only 1 Mbps, the transmission will overload the receiver and will be largely lost.

TCP/IP

- The Internet Protocol (IP) is the most important communications protocol in the Internet protocol suite for relaying data across network boundaries.

- It essentially establishes the Internet.

- In the past, IP did not provide the connectivity; It only specified how packets are supposed to be created.

- The Transmission Control Protocol (TCP) allowed this functionality. Since one could not perform its task without the other, they earned the name TCP/IP to show how they depend on each other.

TCP/IP

☐ Think of IP as something like the postal system. It allows you to address a package and drop it into the system, but there is no actual direct link between you and the recipient.

☐ Instead, there is a "web" of links interconnecting with each other. This is where IP and TCP come in.

☐ IP tells packets what their destination is and how to get there; TCP ensures a reliable connection, checking packets for errors, requesting a "re-transmission" if it detects one.

☐ The Internet Protocol gets information from a source computer to a destination computer. It sends this information in the form of packets.

☐ IP packets or datagram has two parts. The first part is the header, which is like a label on an envelope. The second part is the payload, which is like the letter inside an envelope.

☐ The header contains the source and destination IP addresses, and some extra information. This information is called metadata, and is about the packet itself. Putting data in a packet with a header is encapsulation.

File Transfer Protocol

- FTP, also known as File Transfer Protocol, is a communication protocol for the rapid, simple transmission of computer files across a network supporting the TCP/IP.

- FTP uses the Client-Server architecture, meaning that there is a server, that holds the files, and does the authentication, and a client, or the end-user, who is accessing the files.

- The server listens on the network for connection requests from other computers.

- The client can make a connection to the FTP server by using FTP client software.

- Once connected and authenticated (via rsh or SFTP) the client can do things such as uploading files to the server, downloading files (taking the server's files and putting them on his own computer) from the server, and renaming, deleting files on the server, changing file permissions, etc.

- By using FTP the internet becomes a huge disk drive attached to your computer.

□ FTP sites contain book, journals, software's, games, images, sound, multimedia courseware, databases and other information available for you to transfer.

□ FTP uses Archie server. An agent who uses normal FTP commands to get directory listing of all the files on hundrends of anonymous sites around the world.

Hypertext Transfer Protocol (HTTP)

☐ Hypertext Transfer Protocol(HTTP) is an internet communication protocol used to send and receive webpages and files on the internet.

☐ HTTP defines how messages are <u>formatted</u> and <u>transmitted,</u> and what action web browsers should take in response to various commands.

☐ for e.g., when the user enters a URL in a browser, the browser sends and HTTP command to the web server directing it to fetch and transmit the requested web page.

☐ HTTP is a connectionless text-based protocol. Clients (web browsers) send requests to web servers for web elements such as web pages and images.

☐ After the request is serviced by a server, the connection between client and server across the Internet is disconnected.

☐ A new connection must be made for each request.

- The server must be located using a URL or URI. This always contains http:// at the start.

- It normally connects to port 80 on a computer.

- A more secure version of HTTP is called HTTPS (Hypertext Transfer Protocol Secure).

- This contains https:// at the beginning of the URL. It encrypts all the information that is sent and received.

- This can stop malicious users such as hackers from stealing the information and is often used on payment websites.

- HTTPS uses port 443 for communication instead of port 80.

Simple Mail Transfer Protocol (SMTP)

- Simple Mail Transfer Protocol is a TCP/IP used in sending and receiving e-mail.

- It provides a mail exchange between users on the same or different computers, and it also supports:
  - It can send a single message to one or more recipients.
  - Sending message can include text, voice, video or graphics.
  - It can also send the messages on networks outside the internet.

- The main purpose of SMTP is used to set up communication rules between servers.

- The servers also handles the errors such as incorrect email address.

- For example, if the recipient address is wrong, then receiving server reply with an error message of some kind.
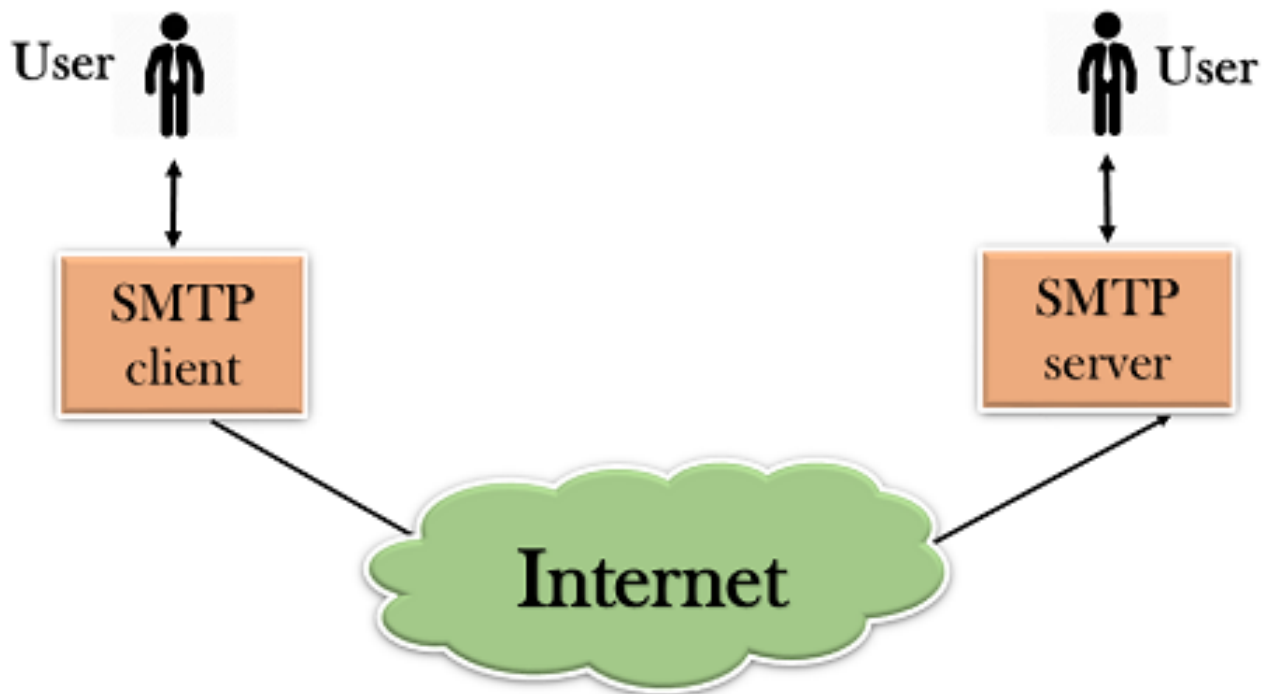
- The SMTP client and server into two components such as User Agent (UA) and Mail Transfer Agent (MTA).

- The User Agent (UA) prepares the message, creates the envelope and then puts the message in the envelope.

- The Mail Transfer Agent (MTA) transfers this mail across the internet.
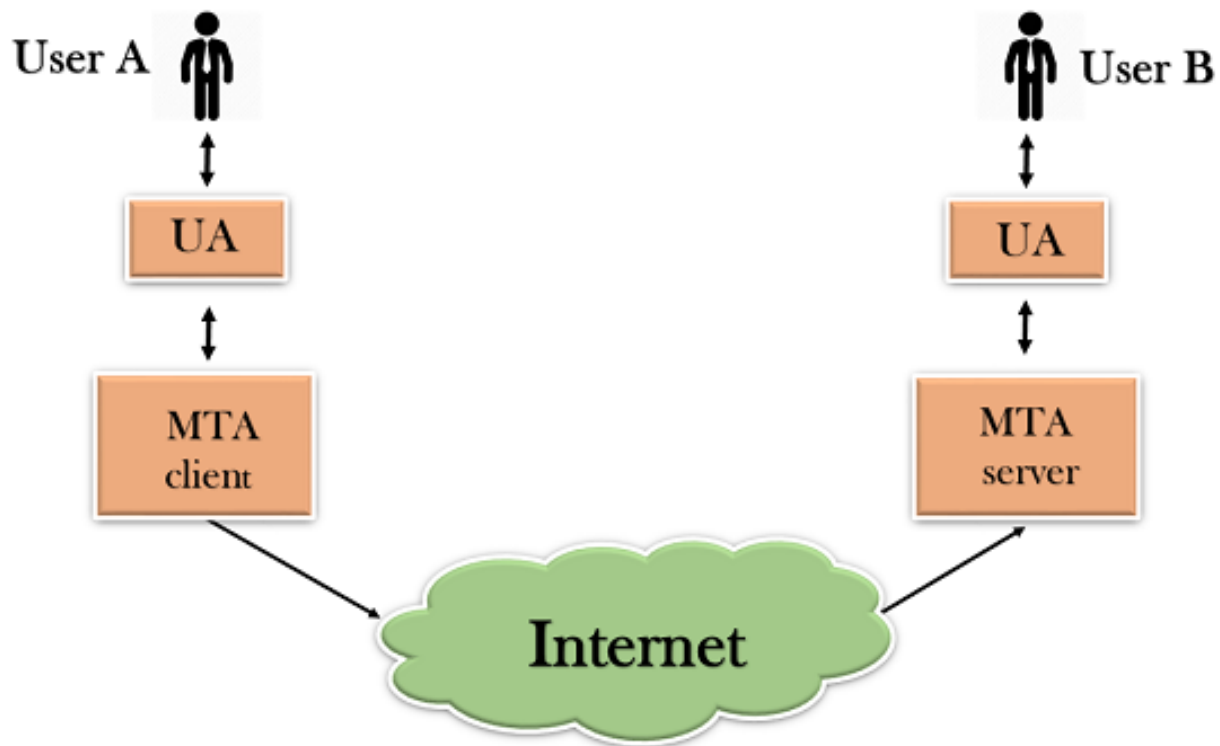
Post Office Protocol

- The POP protocol stands for Post Office Protocol.

- Usually, SMTP is used as a message transfer agent. When the message is sent, then SMTP is used to deliver the message from the client to the server and then to the recipient server. But the message is sent from the recipient server to the actual server with the help of the Message Access Agent.

- The Message Access Agent contains two types of protocols, i.e., POP3 and IMAP.

- The transmission of mail from the sender to the sender's mail server and then to the receiver's mail server is done with the help of the SMTP protocol.

- At the receiver's mail server, the POP or IMAP protocol takes the data and transmits to the actual user.

- Since SMTP is a push protocol so it pushes the message from the client to the server.

- The next stage of email communication requires a pull protocol, and POP is a pull protocol. When the mail is transmitted from the recipient mail server to the client which means that the client is pulling the mail from the server.
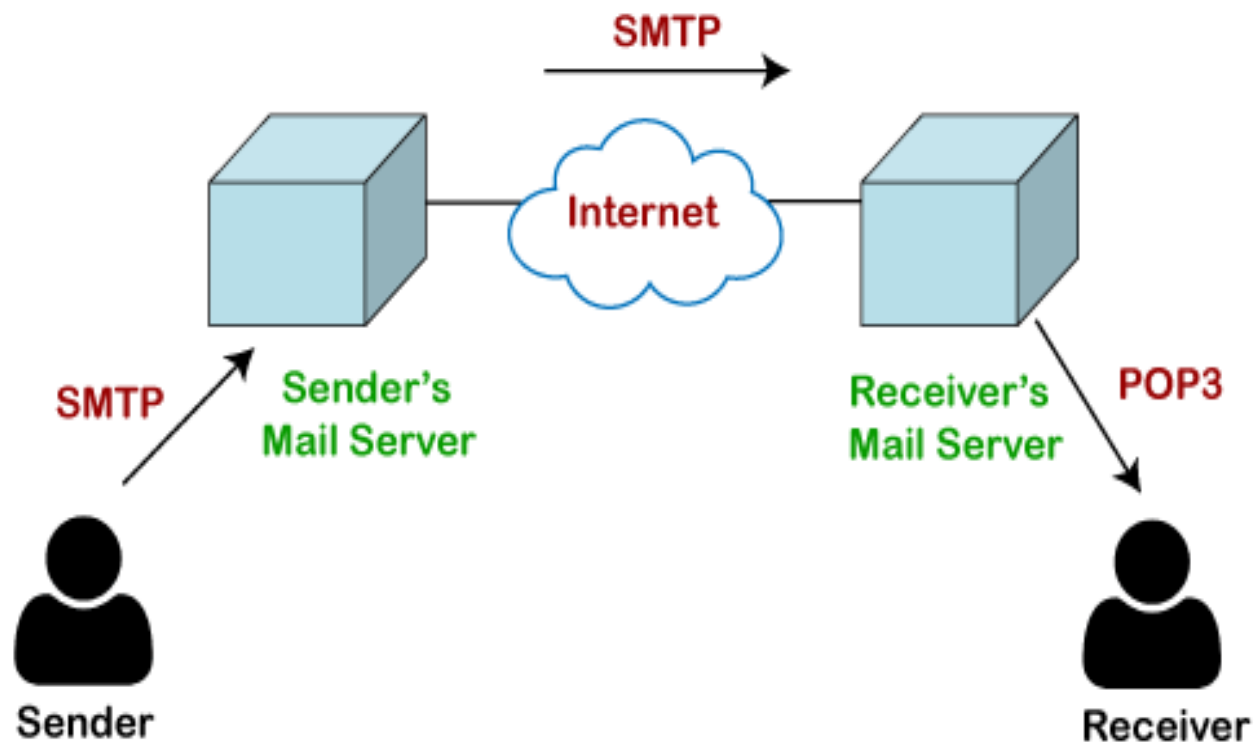
POP3

- Post Office Protocol version 3 (POP3) is a standard mail protocol used to receive emails from a remote server to a local email client.

- POP3 allows you to download email messages on your local computer and read them even when you are offline.

- When we use POP3 to connect to email account, messages are downloaded locally and removed from the email server. This means that accessing an account from multiple locations, that may not be the best option.

- On the other hand, while using POP3, the messages are stored on local computer, which reduces the space email account.

- By default, the POP3 protocol works on two ports:

  - Port 110 – this is the default POP3 non-encrypted port;

  - Port 995 – this is the port you need to use if you want to connect using POP3 securely.

IMAP

- The Internet Message Access Protocol (IMAP) is a mail protocol used for accessing email on a remote web server from a local client. IMAP and POP3 are the two most commonly used Internet mail protocols for retrieving emails. Both protocols are supported by all modern email clients and web servers.

- While the POP3 protocol assumes that your email is being accessed only from one application, IMAP allows simultaneous access by multiple clients. This is why IMAP is more suitable to access email from different locations or messages are managed by multiple users.

- By default, the IMAP protocol works on two ports:

  - Port 143 – this is the default IMAP non-encrypted port;

  - Port 993 – this is the port you need to use if you want to connect using IMAP securely.

Post Office Protocol (POP3)	Internet Message Access Protocol (IMAP)
POP is a simple protocol that only allows downloading messages from your Inbox to your local computer.	IMAP is much more advanced and allows you the user to see all the folders on the mail server.
The POP server listens on port 110, and the POP with SSL secure(POP3DS) server listens on port 995	he IMAP server listens on port 143, and the IMAP with SSL secure(IMAPDS) server listens on port 993.
In POP3 the mail can only be accessed from a single device at a time.	Messages can be accessed across multiple devices
To read the mail it has to be downloaded on the local system.	The mail content can be read partially before downloading.
The user can not organize mails in the mailbox of the mail server.	The user can organize the emails directly on the mail server.
The user can not create, delete or rename email on the mail server.	The user can create, delete or rename email on the mail server.
It has two modes : delete mode and keep mode. In delete mode , the mail is deleted from mail box after retrieval. In keep mode , the mail remains in the mail box after retrieval.	Multiple redundant copies of the message are kept at the mail server, in case of loss of message of a local server, the mail can still be retrieved
Changes in the mail can be done using local email software.	Changes made web interface or email software stay in sync with the server.
All the message are downloaded at once.	Message header can be viewed prior to downloading.

Multipurpose Internet Mail Extension

- Multipurpose Internet Mail Extension (MIME) is a standard that was proposed by Bell Communications in 1991 in order to expand the limited capabilities of email.

- MIME is a kind of addon or a supplementary protocol that allows non-ASCII data to be sent through SMTP. It allows the users to exchange different kinds of data files on the Internet: audio, video, images, application programs as well.

- Email messages with MIME formatting are typically transmitted with standard protocols, such as the Simple Mail Transfer Protocol (SMTP), the Post Office Protocol (POP), and the Internet Message Access Protocol (IMAP)

Need for MIME

- Limitations of Simple Mail Transfer Protocol (SMTP):
  - SMTP has a very simple structure
  - Its simplicity however comes with a price as it only sends messages in Network Virtual Terminal (NVT) 7-bit ASCII format.
  - It cannot be used for languages that do not support 7-bit ASCII format such as- French, German, Russian, Chinese and Japanese, etc. so it cannot be transmitted using SMTP. So, in order to make SMTP more broad, we use MIME.
  - It cannot be used to send binary files or video or audio data.

Purpose and Functionality of MIME

☐ Growing demand for Email Messages as people also want to express themselves in terms of Multimedia. So, MIME another email application is introduced as it is not restricted to textual data.

☐ MIME transforms non-ASCII data at the sender side to NVT 7-bit data and delivers it to the client SMTP. The message on the receiver side is transferred back to the original data. As well as we can send video and audio data using MIME as it transfers them also in 7-bit ASCII data.

Features of MIME

□ It allows sending multiple attachments with a single message.

□ Unlimited message length.

□ Binary attachments (executables, images, audio, or video files) may be divided if needed.

□ MIME provided support for varying content types and multi-part messages.

Working of MIME

□ Suppose a user wants to send an email through a user agent and it is in a non-ASCII format so there is a MIME protocol that converts it into 7-bit NVT ASCII format. The message is transferred through the e-mail system to the other side in the 7-bit format now MIME protocol again converts it back into non-ASCII code and now the user agent of the receiver side reads it and then information is finally read by the receiver. MIME header is basically inserted at the beginning of any e-mail transfer.

MIME with SMTP and POP

- SMTP transfers the mail being a message transfer agent from the sender's side to the mailbox of the receiver side and stores it and the MIME header is added to the original header and provides additional information. while POP being the message access agent organizes the mails from the mail server to the receiver's computer. POP allows user agent to connect with the message transfer agent.

MIME Header:

- It is added to the original e-mail header section to define transformation. There are five headers which we add to the original header:

  - **MIME Version –** Defines version of MIME protocol. It must have the parameter Value 1.0, which indicates that message is formatted using MIME. (Eg: MIME-Version: 1.0)

  - **Content Type –** Type of data used in the body of message. They are of different types like text data (plain, HTML), audio content or video content. (Eg: Content-Type: text/plain, image/jpeg, audio/mp3, video/mp4, and application/msword and so on)

  - **Content Type Encoding –** It defines the method used for encoding the message. Like 7-bit encoding, 8-bit encoding, etc., (Eg: Q-encoded or base64-encoded text).

  - **Content Id –** It is used for uniquely identifying the message.

  - **Content-Disposition:** Content-Disposition headers are included to allow a message to be tagged to indicate its presentation mechanism or its archival disposition, a file name.

  - (Eg: Content-Disposition: inline/attachment; filename="General Quiz.png")

Web Security

- Web application security is a branch of information security that deals specifically with security of websites, web applications and web services. At a high level, web application security draws on the principles of application security but applies them specifically to internet and web systems.

- The global nature of the Internet exposes web properties to attack from different locations and various levels of scale and complexity. Web application security deals specifically with the security surrounding websites, web applications and web services such as APIs.

Web Security Vulnerabilities

☐ Let's explore some of the common methods of attack or "vectors" commonly exploited.

Cross site scripting (XSS) - XSS is a vulnerability that allows an attacker to inject client-side scripts into a webpage in order to access important information directly, impersonate the user, or trick the user into revealing important information.

SQL injection (SQi) - SQi is a method by which an attacker exploits vulnerabilities in the way a database executes search queries. Attackers use SQi to gain access to unauthorized information, modify or create new user permissions, or otherwise manipulate or destroy sensitive data.

□ **Denial-of-service (DoS)** and distributed denial-of-service (DDoS) attacks - Through a variety of vectors, attackers are able to overload a targeted server or its surrounding infrastructure with different types of attack traffic. When a server is no longer able to effectively process incoming requests, it begins to behave sluggishly and eventually deny service to incoming requests from legitimate users.

- **Memory corruption –** Memory corruption occurs when a location in memory is unintentionally modified, resulting in the potential for unexpected behavior in the software. Bad actors will attempt to sniff out and exploit memory corruption through exploits such as code injections or buffer overflow attacks.

- **Buffer overflow –** Buffer overflow is an anomaly that occurs when software writing data to a defined space in memory known as a buffer. Overflowing the buffer's capacity results in adjacent memory locations being overwritten with data. This behavior can be exploited to inject malicious code into memory, potentially creating a vulnerability in the targeted machine.

☐ **Cross-site request forgery (CSRF)** - Cross site request forgery involves tricking a victim into making a request that utilizes their authentication or authorization. By leveraging the account privileges of a user, an attacker is able to send a request masquerading as the user. Once a user's account has been compromised, the attacker can exfiltrate, destroy or modify important information. Highly privileged accounts such as administrators or executives are commonly targeted.

☐ **Data breach -** A data breach is the intentional or unintentional release of secure or private/confidential information to an untrusted environment. Data breaches may involve personal health information (PHI), personally identifiable information (PII), trade secrets or intellectual property.

HTML & XHTML

Dr P.V. Praveen Sundar
Assistant Professor,
Department of Computer Science
Adhiparasakthi College of Arts & Science,
Kalavai.

12-Feb-2021

I MCA

- XHTML stands for **EXtensible HyperText Markup Language**. It is the next step in the evolution of the internet.

- The XHTML 1.0 is the first document type in the XHTML family.

- XHTML is almost identical to **HTML 4.01** with only few differences. This is a cleaner and stricter version of HTML 4.01. If you already know HTML, then you need to give little attention to learn this latest version of HTML.

- XHTML was developed by **World Wide Web Consortium (W3C)** to help web developers make the transition from HTML to XML. By migrating to XHTML today, web developers can enter the XML world with all of its benefits, while still remaining confident in the backward and future compatibility of the content.

Why Use XHTML?

- Developers who migrate their content to XHTML 1.0 get the following benefits −
  - XML is a markup language where all documents must be marked up correctly (be "well-formed").
  - XHTML was developed to make HTML more extensible and flexible to work with other data formats (such as XML). In addition, browsers ignore errors in HTML pages, and try to display the website even if it has some errors in the markup. So XHTML comes with a much stricter error handling.
  - XHTML documents are XML conforming as they are readily viewed, edited, and validated with standard XML tools.
  - XHTML documents can utilize applications such as scripts and applets that rely upon either the HTML Document Object Model or the XML Document Object Model.
  - XHTML gives you a more consistent, well-structured format so that your webpages can be easily parsed and processed by present and future web browsers.
  - Since XHTML is an official standard of the W3C, your website becomes more compatible with many browsers and it is rendered more accurately.
  - XHTML combines strength of HTML and XML. Also, XHTML pages can be rendered by all XML enabled browsers.
  - XHTML defines quality standard for your webpages and if you follow that, then your web pages are counted as quality web pages. The W3C certifies those pages with their quality stamp.

Main differences from HTML

- <!DOCTYPE> is mandatory

- The xmlns attribute in <html> is mandatory

- <html>, <head>, <title>, and <body> are mandatory

- <title> must come first in <head> element.

- Elements must always be properly nested

- Elements must always be closed

- Elements must always be in lowercase

- Attribute names must always be in lowercase

- Attribute values must always be quoted

- Attribute minimization is forbidden

- An unused tags must be minimized.

- Browsers ignore Unknown Attributes and Elements.
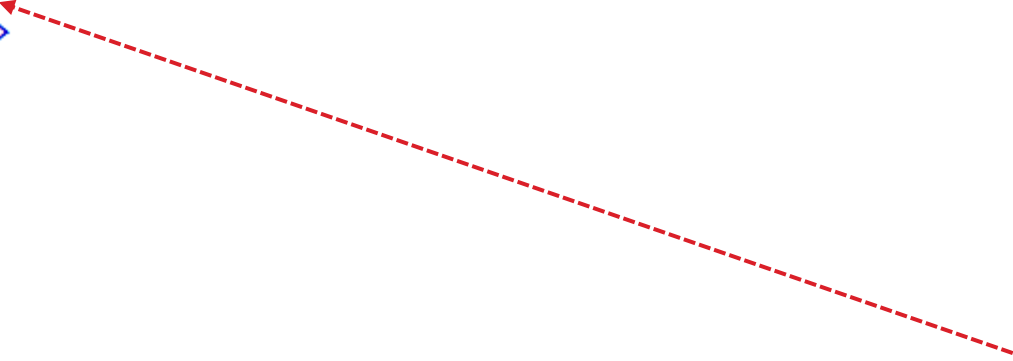
XHTML - <!DOCTYPE ....> Is Mandatory

- An XHTML document must have an XHTML <!DOCTYPE> declaration.

- The <!DOCTYPE> statement indicates the particular version of HTML or XHTML being used in the document.

- The <html>, <head>, <title>, and <body> elements must also be present, and the xmlns attribute in <html> must specify the xml namespace for the document. Notice that under XHTML, it is mandatory to specify a little more information about the language that we are using.

- DTD stands for Document Type Definition. **A DTD defines the structure and the legal elements and attributes of an XML document.**

- **Namespaces are used primarily to avoid conflicts between element names when mixing XML languages. The xmlns attribute specifies the xml namespace for a document.** Note: The xmlns attribute is required in XHTML, invalid in HTML 4.01, and optional in HTML5

Example

Here is an XHTML document with a minimum of required tags:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
"http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
  <title>Title of document</title>
</head>
<body>

  some content here...

</body>
</html>
```

XHTML Elements Must be Properly Nested

In XHTML, elements must always be properly nested within each other, like this:

Correct:

```
<b><i>Some text</i></b>
```

Wrong:

```
<b><i>Some text</b></i>
```

XHTML Elements Must Always be Closed

In XHTML, elements must always be closed, like this:

Correct:

```
<p>This is a paragraph</p>
<p>This is another paragraph</p>
```

Wrong:

```
<p>This is a paragraph
<p>This is another paragraph
```

In XHTML, empty elements must always be closed, like this:

Correct:

```
A break: <br />
A horizontal rule: <hr />
An image: <img src="happy.gif" alt="Happy face" />
```

Wrong:

```
A break: <br>
A horizontal rule: <hr>
An image: <img src="happy.gif" alt="Happy face">
```

XHTML Elements Must be in Lowercase

In XHTML, element names must always be in lowercase, like this:

Correct:

```
<body>
<p>This is a paragraph</p>
</body>
```

Wrong:

```
<BODY>
<P>This is a paragraph</P>
</BODY>
```

XHTML Attribute Names Must be in Lowercase

In XHTML, attribute names must always be in lowercase, like this:

Correct:

```
<a href="https://www.apcas.in/"> Adhiparasakthi college of Arts and Science</a>
```

Wrong:

```
<a HREF="https://www.apcas.in/"> Adhiparasakthi college of Arts and Science</a>
```

9

XHTML Attribute Values Must be Quoted

In XHTML, attribute values must always be quoted, like this:

Correct:

```
<a href="https://www.apcas.in"> Adhiparasakthi College of Arts and Science </a>
```

Wrong:

```
<a href=https://www.apcas.in> Adhiparasakthi College of Arts and Science </a>
```

XHTML Attribute Minimization is Forbidden

In XHTML, attribute minimization is forbidden:

Correct:

```
<input type="checkbox" name="vehicle" value="car" checked="checked" />
<input type="text" name="lastname" disabled="disabled" />
```

Wrong:

```
<input type="checkbox" name="vehicle" value="car" checked />
<input type="text" name="lastname" disabled />
```

HTML Style	XHTML Style
compact	compact="compact"
checked	checked="checked"
declare	declare="declare"
readonly	readonly="readonly"
disabled	disabled="disabled"
selected	selected="selected"
defer	defer="defer"
ismap	ismap="ismap"
nohref	nohref="nohref"
noshade	noshade="noshade"
nowrap	nowrap="nowrap"
multiple	multiple="multiple"
noresize	noresize="noresize"

11

The *id* Attribute

The id attribute replaces the name attribute. Instead of using name = "name", XHTML prefers to use id = "id". The following example shows how −

```
<!-- This is invalid in XHTML -->
<img src="/images/xhtml.gif" name="xhtml_logo" />


<!-- Correct XHTML way of writing this is as follows -->
<img src="/images/xhtml.gif" id="xhtml_logo" />
```

The *language* Attribute

The language attribute of the script tag is deprecated. The following example shows this difference −

```
<!-- This is invalid in XHTML -->

<script language="JavaScript" type="text/JavaScript">
    document.write("Hello XHTML!");
</script>


<!-- Correct XHTML way of writing this is as follows -->

<script type="text/JavaScript">
    document.write("Hello XHTML!");
</script>
```

12

S.No.	HTML	XHTML
1.	HTML stands for Hypertext Markup Language.	XHTML stands for Extensible Hypertext Markup Language.
2.	It was developed by Tim Berners-Lee.	It was developed by W3C i.e World Wide Web Consortium.
3.	It was developed in 1991.	It was released in 2000.
4.	It is extended from SGML.	It is extended from XML and HTML.
5.	The format is a document file format.	The format is a markup language.
6.	All tags and attributes are not necessarily to be in lower or upper case.	In this, every tag and attribute should be in lower case.
7.	Doctype is not necessary to write at the top.	Doctype is very necessary to write at the top of the file.
8.	It is not necessary to close the tags in the order they are opened.	It is necessary to close the tags in the order they are opened.
9.	While using the attributes it is not necessary to mention quotes. For e.g. <apcas>.	While using the attributes it is mandatory to mention quotes. For e.g. <id="Apcas">.
10.	Filename extension used are .html, .htm.	Filename extension are .xhtml, .xht, .xml.

Welcome to the World of HTML

HTML **really** isn't so hard!

HTML is a simplest markup language, which is used to create a webpages.

First_HTML - Notepad

File Edit Format View Help

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01
Transitional// EN"
"http://www.w3.org/TR/html4/loose.dtd">
<html>
<head>
<title> First HTML Example </title>
</head>
<body>
<h1> Welcome to the World of HTML </h1>
<hr>
<p> HTML <b> really </b> isn't so hard! </p>
<p> HTML is a simplest markup language, which is used
to create a webpages.</p>
```

Ln 11, Col 1   100%   Windows (CRLF)   UTF-8
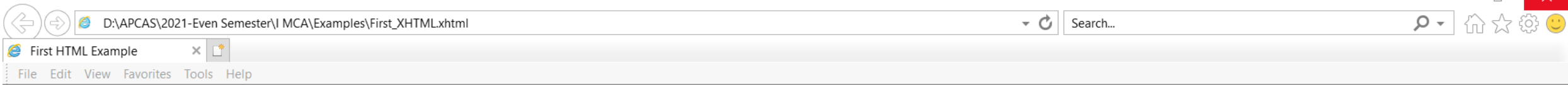
Welcome to the World of HTML

HTML **really** isn't so hard!

HTML is a simplest markup language, which is used to create a webpages.

First_XHTML - Notepad

File Edit Format View Help

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0
Transitional// EN"
"http://www.w3.org/TR/xhtml11-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title> First HTML Example </title>
</head>
<body>
<h1> Welcome to the World of HTML </h1>
<hr/>
<p> HTML <b> really </b> isn't so hard! </p>
<p> HTML is a simplest markup language, which is used
to create a webpages.</p>
</body>
</html>
```

Ln 12, Col 8   100%   Windows (CRLF)   UTF-8

XML Parsing Error: syntax error
Location: file:///D:/APCAS/2021-Even%20Semester/I%20MCA/Examples/First_XHTML.xhtml
Line Number 1, Column 1:

```
<!doctype HTML PUBLIC "-//W3C//DTD XHTML 1.0 Transitional// EN" "http://www.w3.org/TR/xhtml11-transitional.dtd">
^
```

First_XHTML - Notepad

File   Edit   Format   View   Help

```
<!doctype HTML PUBLIC "-//W3C//DTD XHTML 1.0
Transitional// EN" "http://www.w3.org/TR/xhtml11-
transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title> First HTML Example </title>
</head>
<body>
<h1> Welcome to the World of HTML </h1>
<hr/>
<p> HTML <b> really </b> isn't so hard! </p>
<p> HTML is a simplest markup language, which is used
to create a webpages.</p>
</body>
</html>
```

First HTML Example   ×   http://www.w3.org/TR/xhtml11   ×   W3 W3C QA - Recommended list   ×   W3 google.co.in/url?sa=t&rct=j&q   ×   First HTML Example   ×   +

file:///D:/APCAS/2021-Even Semester/I MCA/Examples/First_XHTML.xhtml   ...   ⊙   ☆   Search

XML Parsing Error: no root element found
Location: file:///D:/APCAS/2021-Even%20Semester/I%20MCA/Examples/First_XHTML.xhtml
Line Number 12, Column 1:

^

First_XHTML - Notepad   —   □   ×

File   Edit   Format   View   Help

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0
Transitional// EN" "http://www.w3.org/TR/xhtml11-
transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title> First HTML Example </title>
</head>
<body>
<h1> Welcome to the World of HTML </h1>
<hr/>
<p> HTML <b> really </b> isn't so hard! </p>
<p> HTML is a simplest markup language, which is used
to create a webpages.</p>
```

Type here to search   O   ⊞   L   ✉   🦊   P   📘   ℮   04:18 AM   11-02-2021

First HTML Example   ✕   http://www.w3.org/TR/xhtml11 ✕   W3C QA - Recommended list ✕   google.co.in/url?sa=t&rct=j&q ✕   First HTML Example   ✕   +

file:///D:/APCAS/2021-Even Semester/I MCA/Examples/First_XHTML.xhtml   ...   🔍 Search

XML Parsing Error: mismatched tag. Expected: </hr>.
Location: file:///D:/APCAS/2021-Even%20Semester/I%20MCA/Examples/First_XHTML.xhtml
Line Number 11, Column 3:

```
</body>
--^
```

First_XHTML - Notepad

File   Edit   Format   View   Help

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0
Transitional// EN" "http://www.w3.org/TR/xhtml11-
transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title> First HTML Example </title>
</head>
<body>
<h1> Welcome to the World of HTML </h1>
<hr>
<p> HTML <b> really </b> isn't so hard! </p>
<p> HTML is a simplest markup language, which is used
to create a webpages.</p>
</body>
</html>
```

Type here to search   O   📧   ENG   04:19 AM   11-02-2021

First HTML Example | http://www.w3.org/TR/xhtml11 | W3C QA - Recommended list o | google.co.in/url?sa=t&rct=j&q | First HTML Example

file:///D:/APCAS/2021-Even Semester/I MCA/Examples/First_XHTML.xhtml

Search

XML Parsing Error: mismatched tag. Expected: </html>.
Location: file:///D:/APCAS/2021-Even%20Semester/I%20MCA/Examples/First_XHTML.xhtml
Line Number 12, Column 3:

</HTML>
--^

First_XHTML - Notepad

File   Edit   Format   View   Help

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0
Transitional// EN" "http://www.w3.org/TR/xhtml11-
transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title> First HTML Example </title>
</head>
<body>
<h1> Welcome to the World of HTML </h1>
<hr/>
<p> HTML <b> really </b> isn't so hard! </p>
<p> HTML is a simplest markup language, which is used
to create a webpages.</p>
</body>
</HTML>
```

Type here to search

04:20 AM
ENG
11-02-2021

Welcome to the World of HTML

HTML **really** isn't so hard!

HTML is a simplest markup language, which is used to create a webpages.

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0
Transitional// EN" "http://www.w3.org/TR/xhtml11-
transitional.dtd">
<HTML xmlns="http://www.w3.org/1999/xhtml" lang="en">
<title> First HTML Example </title>
<body>
<h1> Welcome to the World of HTML </h1>
<hr/>
<p> HTML <b> really </b> isn't so hard! </p>
<p> HTML is a simplest markup language, which is used
to create a webpages.</p>
</body>
</HTML>
```

HTML5 and beyond

```
<!DOCTYPE HTML>
```

HTML 4.01
  Strict

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
    "http://www.w3.org/TR/html4/strict.dtd">
```

  Transitional

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
    "http://www.w3.org/TR/html4/loose.dtd">
```

  Frameset

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Frameset//EN"
    "http://www.w3.org/TR/html4/frameset.dtd">
```

XHTML 1.0
  Strict (**quick reference**)

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
```

  Transitional

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
```

  Frameset

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Frameset//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-frameset.dtd">
```

XHTML 1.1 - **DTD**:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
    "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
```

XHTML Basic 1.1 (**quick reference**):

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML Basic 1.1//EN"
    "http://www.w3.org/TR/xhtml-basic/xhtml-basic11.dtd">
```

HTML

Living Standard — Last Updated 10 February 2021

Table of contents

File an issue about the selected text

Basic Structure of XHTML

- Every XHTML document has three major components: the XHTML declaration, the head, and the body. The document structure elements are those that define each component.

- An XHTML 1.0 document begins with a DOCTYPE declaration that declares the version of XHTML to which the document conforms.

- <html> tag declares the document to be an XHTML document. All document content and supporting XHTML code goes between the <html> and </html> elements.

- The <html> element can take the xmlns attribute, which specifies the URL that describes the XML namespace for the document (for example, xmlns="http://www.w3.org/1999/xhtml").

- The html element follows and contains the head and body.

- The head contains information about the document, such as its title and keywords, while the body contains the actual content of the document, made up of block-level elements and inline elements.

General form of <HTML> tag

```
<html
        dir="ltr | rtl"
        id="unique alphanumeric identifier"
        lang="language code"
        xmlns="http://www.w3.org/1999/xhtml>
</html>
```

- The html element is the first element in a document. Except for comments. It contains <head> and followed by <body> or <frameset>

- The <html> tag and its closing tag </html> are both mandatory under XHTML.

Minimal XHTML Document

```
<!DOCTYPE .....>
  <html>
      <head>
          .....
      </head>


      <body>
          <p>
          .......
          </p>
      </body>
  </html>
```

HTML element

HTML element

HTML element

HTML element

HTML element
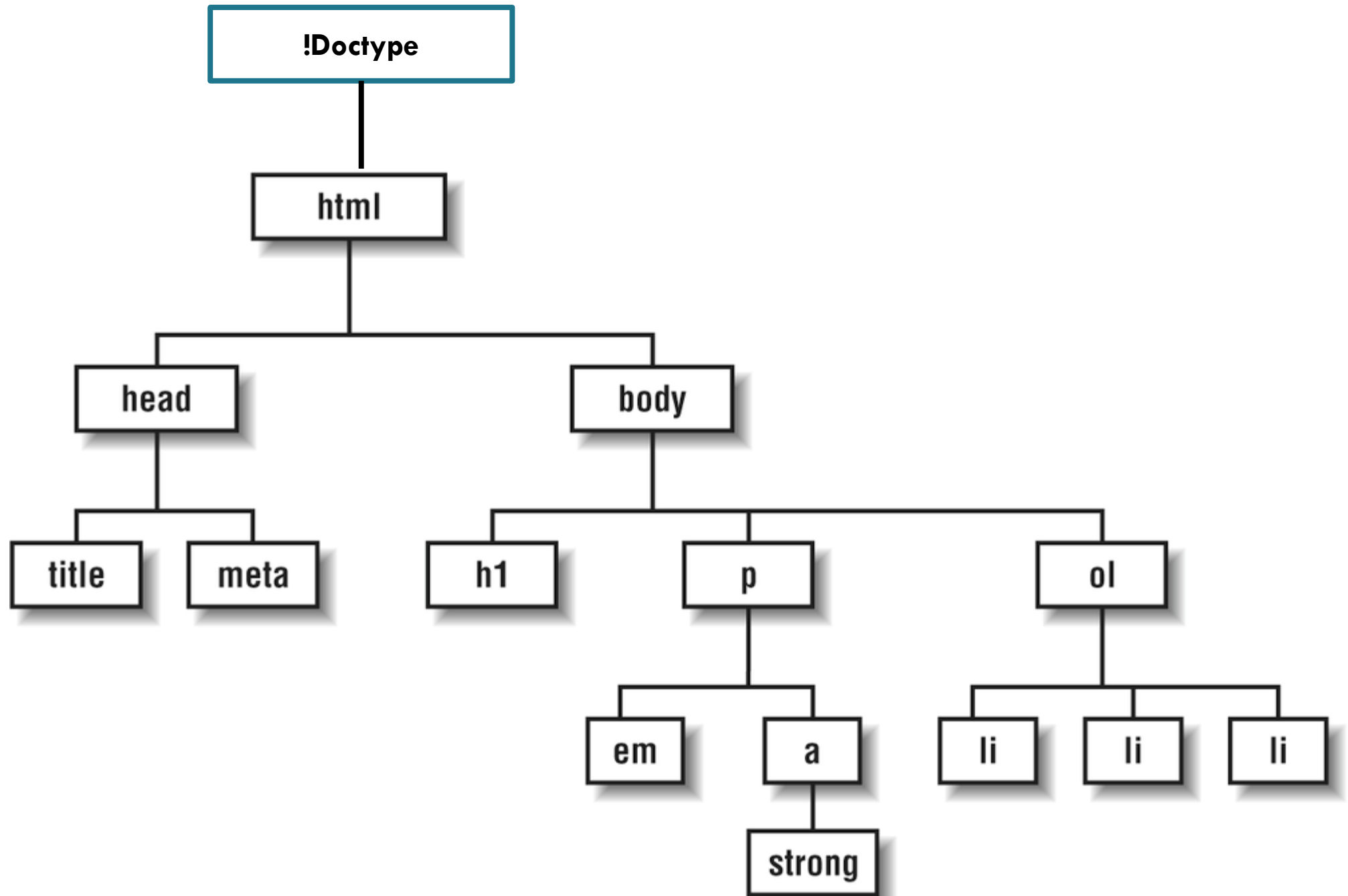
XHTML - Doctypes

- The XHTML standard defines three Document Type Definitions (DTDs). The most commonly used and easy one is the XHTML Transitional document.

- XHTML 1.0 document type definitions correspond to three DTDs −
  - Strict
  - Transitional
  - Frameset

- There are few XHTML elements and attributes, which are available in one DTD but not available in another DTD. Therefore, while writing your XHTML document, you must select your XHTML elements or attributes carefully. However, XHTML validator helps you to identify valid and invalid elements and attributes.

- The doctype statement should be used as the first line of all documents.

- Validation program might use this construct when determining the correctness of an html document.

- Modern browsers may determine what rendering mode to use depending on the doctype statement.

- An incorrect doctype that does not correspond to appropriate markup usuage may result in inaccurate display.

- Examples

<! DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 TRANSITIONAL//EN"

<! DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0 TRANSITIONAL//EN"

<! DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1 //EN" "xhtml111.dtd"

<! DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

HTML Version	!DOCTYPE Declaration
2.0	<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML//EN">
3.2	<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2 Final//EN">
4.0 Transitional	<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">
4.0 Frameset	<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Frameset//EN" "http://www.w3.org/TR/html4/frameset.dtd">
4.0 Strict	<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">
4.01 Transitional	<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">
4.01 Frameset	<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Frameset//EN" "http://www.w3.org/TR/html4/frameset.dtd">
4.01 Strict	<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">
XHTML 1.0 Transitional	<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
XHTML 1.0 Strict	<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
XHTML 1.0 Frameset	<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Frameset//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-frameset.dtd">
XHTML 1.1	<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN" "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
XHTML 2.0 (still in progress)	<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 2.0//EN" " http://www.w3.org/TR/xhtml2/DTD/xhtml2.dtd ">

TABLE 1-1   Common HTML Doctype Declarations

HTML Version	Description
2.0	Classic HTML dialect supported by browsers such as Mosaic. This form of HTML supports core HTML elements and features such as tables and forms but does not consider any of the browser innovations of advanced features such as style sheets, scripting, or frames.
3.0	The proposed replacement for HTML 2.0 that was never widely adopted, most likely due to the heavy use of browser-specific markup.
3.2	An HTML finalized by the W3C in early 1997 that standardized most of the HTML features introduced in browsers such as Netscape 3. This version of HTML supports many presentation elements, such as fonts, as well as early support for some scripting features.
4.0 Transitional	The 4.0 transitional form finalized by the W3C in December of 1997 preserves most of the presentation elements of HTML 3.2. It provides a basis for transition to CSS as well as a base set of elements and attributes for multiple language support, accessibility, and scripting.
4.0 Strict	The strict version of HTML 4.0 removes most of the presentation elements from the HTML specification, such as fonts, in favor of using Cascading Style Sheets (CSS) for page formatting.
4.0 Frameset	The frameset specification provides a rigorous syntax for framed documents that was lacking in previous versions of HTML.
4.01 Transitional/ Strict/Frameset	A minor update to the 4.0 standard that corrects some of the errors in the original specification.
XHTML 1.0 Transitional	A reformulation of HTML as an XML application. The transitional form preserves many of the basic presentation features of HTML 4.0 transitional but applies the strict syntax rules of XML to HTML.
XHTML 1.0 Strict	A reformulation of HTML 4.0 strict using XML. This language is rule enforcing and leaves all presentation duties to technologies such as Cascading Style Sheets (CSS).
XHTML 1.1	A minor change to XHTML 1.0 that restructures the definition of XHTML 1.0 to modularize it for easy extension. It is not commonly used at the time of this writing and only offers minor gains over XHTML 1.0.
XHTML 2.0	A new implementation of XHTML circa 2003 that may not provide backward compatibility with XHTML 1.0 and traditional HTML. XHTML 2 will likely remove most or all presentational tags left in HTML and will introduce even more logical ideas to the language.

TABLE 1-2   Description of Common HTML Versions

XHTML 1.0 Strict

If you are planning to use Cascading Style Sheet (CSS) strictly and avoiding to write most of the XHTML attributes, then it is recommended to use this DTD. A document conforming to this DTD is of the best quality.

If you want to use XHTML 1.0 Strict DTD then you need to include the following line at the top of your XHTML document.

```
/DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
```

XHTML 1.0 Transitional

If you are planning to use many XHTML attributes as well as few Cascading Style Sheet properties, then you should adopt this DTD and you should write your XHTML document accordingly.

If you want to use XHTML 1.0 Transitional DTD, then you need to include the following line at the top of your XHTML document.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.o
```

XHTML 1.0 Frameset

You can use this when you want to use HTML Frames to partition the browser window into two or more frames.

If you want to use XHTML 1.0 Frameset DTD, then you need to include following line at the top of your XHTML document.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Frameset//EN" "http://www.w3.org/T
```

Note − No matter what DTD you are using to write your XHTML document; if it is a valid XHTML document, then your document is considered as a good quality document.

Browser	Features Introduced	Standards Support
Netscape 2.x	Java, JavaScript, Frames, Plug-ins	2.0 and Netscape extensions, many of which became 3.2 standard.
Netscape 3.x	A few proprietary elements such as spacer and multicol	3.2 with Netscape extensions.
Netscape 4.x	Basic CSS support and the proprietary HTML element layer	3.2, part of 4.0, part of CSS1, and Netscape extensions.
Netscape 6.x	Heavy standards support	4.0, CSS1, much of CSS2, good portion of Document Object Model.
Netscape 7.x / Mozilla 1.x	Very heavy standards support, minor CSS extensions, link prefetching	4.0, XHTML 1.0, CSS1 and most of CSS2, nearly all DOM 1 and 2.
Internet Explorer 3.0	Frames and Inline Frames, Jscript, ActiveX controls, VBScript, some proprietary HTML elements such as marquee and bgsound	3.2 with some Microsoft extensions and a limited amount of CSS1.
Internet Explorer 4.0	Significant JavaScript access to page elements	Most of 4.0 with Microsoft extensions, most of CSS1.
Internet Explorer 5.0/5.5/6.0	Native XML support, close to full Document Object Model Level 1	4.0 with Microsoft extensions, most of CSS1, most of the Document Object Model 1. XHTML support in IE 6. Still certainly underlying reliance on IE flavored syntax.
Opera 7.0	Very heavy standards focus, focus on accessibility and speed features and related W3C specifications	4.0 and XHTML, most CSS1 and Document Object Model 1. Pre 7.0 versions of Opera had problems with JavaScript and advanced CSS.
WebTV/MSN TV	Proprietary tags useful for television screen layout and integration with television viewing	3.2 with WebTV/MSN TV extensions. Support for many Netscape and Microsoft extensions.

TABLE 1-3 Web Development Feature Overview By Browser Version

XHTML - Attributes

There are a few XHTML/HTML attributes which are standard and associated to all the XHTML/HTML tags. These attributes are listed here with brief description −

Core Attributes

Not valid in base, head, html, meta, param, script, style, and title elements.

Attribute	Value	Description
class	class_rule or style_rule	The class of the element.
Id	id_name	A unique id for the element.
style	style_definition	An inline style definition.
Title	tooltip_text	A text to display in a mouse tip.

Language Attributes

The *lang* attribute indicates the language being used for the enclosed content. The language is identified using the ISO standard language abbreviations, such as **fr** for French, **en** for English, and so on. More codes and their formats are described at www.ietf.org ☑ .

Not valid in base, br, frame, frameset, hr, iframe, param, and script elements.

Attribute	Value	Description
dir	ltr	rtl	Sets the text direction.
lang	language_code	Sets the language code.

id attribute

- The id attribute is used to a set a unique name for an element in a document.

- For example, using id with the paragraph tag, <p>

  <p id="FirstParagraph"> This is the first paragraph of text. </p>

Names the bound the element "FirstParagraph".

- Naming an element is useful for manipulating the enclosed contents with stylesheet or script.

- For example, a stylesheet rule such as

  <style type="text/css">
  #FirstParagraph {Color:red;}
  </style>
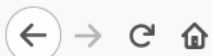
First HTML Example

file:///D:/APCAS/2021-Even Semester/I MCA/Examples/First_XHTML.xhtml

Search

Welcome to the World of HTML

HTML **really** isn't so hard!

HTML is a simplest markup language, which is used to create a webpages.

First_XHTML - Notepad

File   Edit   Format   View   Help

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0
Transitional// EN" "http://www.w3.org/TR/xhtml11-
transitional.dtd">
<HTML xmlns="http://www.w3.org/1999/xhtml" lang="en">
<title> First HTML Example </title>
<body>
<h1> Welcome to the World of HTML </h1>
<hr/>
<p id="First Paragraph"> HTML <b> really </b> isn't
so hard! </p>
<p id="Second Paragraph"> HTML is a simplest markup
language, which is used to create a webpages.</p>

</body>
</HTML>
```

Ln 9, Col 1   100%   Windows (CRLF)   UTF-8

Type here to search   04:22 AM
                                                   ENG   12-02-2021

Class attribute

☐ The class attribute is used to indicate the class or classes that a tagmight belong to. Like id, class is used to associate a tag with a name, so

<p id="First Paragraph" class="important"> HTML really isn't so hard! </p>

☐ The above tag not only indicates the name of the paragraph and also indicates that this paragraph belongs to a class grouping called important.

☐ The main use of the class attribute is to relate a group of elements to various stylesheet rules.

☐ For eg:

<style type="text/css">

.important {background-color:yellow}

</style>

file:///D:/APCAS/2021-Even Semester/I MCA/Examples/First_XHTML.xhtml

Search

Welcome to the World of HTML

HTML **really** isn't so hard!

HTML is a simplest markup language, which is used to create a webpages.

First_XHTML - Notepad

File   Edit   Format   View   Help

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0
Transitional// EN" "http://www.w3.org/TR/xhtml11-
transitional.dtd">
<HTML xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title> First HTML Example </title>
<style type="text/css">
.important {background-color:yellow}
</style>
</head>
<body>
<h1> Welcome to the World of HTML </h1>
<hr/>
<p id="First Paragraph" class="important"> HTML <b>
really </b> isn't so hard! </p>
<p id="Second Paragraph"> HTML is a simplest markup
language, which is used to create a webpages.</p>
</body>
</HTML>
```

Ln 6, Col 36   100%   Windows (CRLF)   UTF-8

Type here to search

04:38 AM
12-02-2021

Style attribute

□ The style attribute is used to add stylesheet information directly to a tag. For example,

<p id="Second Paragraph" style= "font-size:16pt; Color:red;"> HTML is a simplest markup language, which is used to create a webpages.</p>

□ The above attribute sets the font size of the paragraph to be 16pt, red text. Although the style attribute allows CSS rules to be added to an element with ease, it is preferable to use id or class to relate a document-wide or linked stylesheet.

First HTML Example

file:///D:/APCAS/2021-Even Semester/I MCA/Examples/First_XHTML.xhtml

Search

Welcome to the World of HTML

HTML **really** isn't so hard!

HTML is a simplest markup language, which is used to create a webpages.

First_XHTML - Notepad

File   Edit   Format   View   Help

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0
Transitional// EN" "http://www.w3.org/TR/xhtml11-
transitional.dtd">
<HTML xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title> First HTML Example </title>
<style type="text/css">
.important {background-color:yellow}
</style>
</head>
<body>
<h1> Welcome to the World of HTML </h1>
<hr/>
<p id="First Paragraph" class="important"> HTML <b>
really </b> isn't so hard! </p>
<p id="Second Paragraph" style= "font-size:16pt;
Color:red;"> HTML is a simplest markup language, which is
used to create a webpages.</p>
</body>
</HTML>
```

Ln 13, Col 61   100%   Windows (CRLF)   UTF-8

Type here to search

04:45 AM
ENG
12-02-2021

Title attribute

□ The Title is used to provide advisory text about the element or its contents. In the case of

<p id="FirstParagraph" class="important" title="Example of Title attribute"> HTML really isn't so hard! </p>

The title attribute is set to indicate that this particular paragraph is the example of title attribute. Browsers can display this advisory text in the form of a Tooltip.

Welcome to the World of HTML

HTML **really** isn't so hard!

Example of Title attribute

HTML is a simplest markup language, which is used to create a webpages.

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0
Transitional// EN" "http://www.w3.org/TR/xhtml11-
transitional.dtd">
<HTML xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title> First HTML Example </title>
<style type="text/css">
.important {background-color:yellow}
</style>
</head>
<body>
<h1> Welcome to the World of HTML </h1>
<hr/>
<p id="FirstParagraph" class="important" title="Example
of Title attribute"> HTML <b> really </b> isn't so hard!
</p>
<p id="SecondParagraph" style= "font-size:16pt;
Color:red;"> HTML is a simplest markup language, which is
used to create a webpages.</p>
</body>
</HTML>
```

XHTML - Events

- Events enable the scripts to responds to the user interactions and change the page accordingly and respond to a user while moving the mouse, scolling up or down the screen or entering strokes.

- It is possible to write our own event handlers in JavaScript or VBScript and can specify these event handlers as a value of event tag attribute. The XHTML 1.0 has a similar set of events which is available in HTML 4.01 specification.

The <body> and <frameset> Level Events

- There are only two attributes which can be used to trigger any JavaScript or VBScript code, when any event occurs at document level.

Attribute	Value	Description
onload	Script	Script runs when a XHTML document loads.
onunload	Script	Script runs when a XHTML document unloads.

The <form> Level Events

- There are following six attributes which can be used to trigger any JavaScript or VBScript code when any event occurs at form level.

Attribute	Value	Description
onchange	Script	Script executes when the element changes.
onsubmit	Script	Script executes when the form is submitted.
onreset	Script	Script executes when the form is reset.
onselect	Script	Script executes when the element is selected.
onblur	Script	Script executes when the element loses focus.
onfocus	Script	Script runs when the element gets focus.

Keyboard Events

- The following three events are generated by keyboard. These events are not valid in base, bdo, br, frame, frameset, head, html, iframe, meta, param, script, style, and title elements.

Attribute	Value	Description
onkeydown	Script	Script executes on key press.
onkeypress	Script	Script executes on key press and release.
onkeyup	Script	Script executes key release.

Mouse Events

□ The following seven events are generated by mouse when it comes in contact with any HTML tag. These events are not valid in base, bdo, br, frame, frameset, head, html, iframe, meta, param, script, style, and title elements.

Attribute	Value	Description
onclick	Script	Script executes on a mouse click.
ondblclick	Script	Script executes on a mouse double-click.
onmousedown	Script	Script executes when mouse button is pressed.
onmousemove	Script	Script executes when mouse pointer moves.
onmouseout	Script	Script executes when mouse pointer moves out of an element.
onmouseover	Script	Script executes when mouse pointer moves over an element.
onmouseup	Script	Script executes when mouse button is released.

XHTML 1.1 Modules

The XHTML 1.1 document type is made up of the following XHTML modules.

- **Structure Module** − The Structure Module defines the major structural elements for XHTML. These elements effectively act as the basis for the content model of many XHTML family document types. The elements and attributes included in this module are − body, head, html, and title.

- **Text Module** − This module defines all of the basic text container elements, attributes, and their content model − abbr, acronym, address, blockquote, br, cite, code, dfn, div, em, h1, h2, h3, h4, h5, h6, kbd, p, pre, q, samp, span, strong, and var.

- **Hypertext Module** − The Hypertext Module provides the element that is used to define hypertext links to other resources. This module supports element a.

- **List Module** − As its name suggests, the List Module provides list-oriented elements. Specifically, the List Module supports the following elements and attributes − dl, dt, dd, ol, ul, and li.

- **Object Module** − The Object Module provides elements for general-purpose object inclusion. Specifically, the Object Module supports − object and param.

- **Presentation Module −** This module defines elements, attributes, and a minimal content model for simple presentation-related markup − b, big, hr, i, small, sub, sup, and tt.

- **Edit Module −** This module defines elements and attributes for use in editing-related markup − del and ins.

- **Bidirectional Text Module −** The Bi-directional Text module defines an element that can be used to declare the bi-directional rules for the element's content − bdo.

- **Forms Module −** It provides all the form features found in HTML 4.0. Specifically, it supports − button, fieldset, form, input, label, legend, select, optgroup, option, and textarea.

- **Table Module −** It supports the following elements, attributes, and content model − caption, col, colgroup, table, tbody, td, tfoot, th, thead, and tr.

- **Image Module −** It provides basic image embedding and may be used in some implementations of client side image maps independently. It supports the element − img.

- **Client-side Image Map Module** − It provides elements for client side image maps − area and map.

- **Server-side Image Map Module** − It provides support for image-selection and transmission of selection coordinates. The Server-side Image Map Module supports − attribute ismap on img.

- **Intrinsic Events Module** − It supports all the events discussed in XHTML Events.

- **Meta information Module** − The Meta information Module defines an element that describes information within the declarative portion of a document. It includes element meta.

- **Scripting Module** − It defines the elements used to contain information pertaining to executable scripts or the lack of support for executable scripts. Elements and attributes included in this module are − noscript and script.

□ **Style Sheet Module** − It defines an element to be used when declaring internal style sheets. The element and attribute defined by this module is − style.

□ **Style Attribute Module (Deprecated)** − It defines the style attribute.

□ **Link Module** − It defines an element that can be used to define links to external resources. It supports link element.

□ **Base Module** − It defines an element that can be used to define a base URI against which relative URIs in the document are resolved. The element and attribute included in this module is − base.

Basic Text Formatting Tags

- Formatting is the structuring and styling of the document.

- Formatting allows the designer to determine the appearance and create the style of the XHTML document.

- The formatting tags in XHTML are similar in nature to HTML, but in XHTML, extensive formatting is done only through stylesheets.

- Some of the tags that were previously used in HTML have been deprecated in XHTML.

- In HTML and XHTML the formatting tags are divided into two categories:
  - Physical tag: These tags are used to provide the visual appearance to the text.
  - Logical tag: These tags are used to add some logical or semantic value to the text.

Element	Meaning	Notes
<i> ... </i>	Italics	
 ... 	Bold	
<tt> ... </tt>	Teletype (monospaced)	
<u> ... </u>	Underline	Deprecated in HTML and XHTML strict variants.
<s> ... </s>	Strikethrough	Deprecated in HTML and XHTML strict variants.
<strike> ... </strike>	Strikethrough	Deprecated in HTML and XHTML strict variants.
_{...}	Subscript	
^{...}	Superscript	
<big> ... </big>	Bigger font (one size bigger)	
...	Smaller font (one size smaller)	

TABLE 3-1   Inline Physical Tags

Element	Meaning	Common Rendering
<abbr> ... </abbr>	Abbreviation (for example, Mr.)	Plain
<acronym> ... </acronym>	Acronym (for example, WWW)	Plain
<cite> ... </cite>	Citation	Italics
<code> ... </code>	Code listing	Fixed Width
<dfn> ... </dfn>	Definition	Italics
 ... 	Emphasis	Italics
<kbd> .. </kbd>	Keystrokes	Fixed Width
<q> ... </q>	Inline quotation	Quoted (not in IE 6)
<samp> ... </samp>	Sample text (example)	Fixed Width
 ... 	Strong emphasis	Bold
<var> ... </var>	Programming variable	Italics

TABLE 3-2   Logical Text Formatting Elements

Tag	Description
<b>	Bold text
<strong>	Important text
<i>	Italic text
<u>	Underline Text
<em>	Emphasized text
<mark>	Highlighted text
<small>	Smaller text
<del>	Deleted text
<ins>	Inserted text
<sub>	Subscript text
<sup>	Superscript text
<strike> and <s>	Strikethrough the text with a thin line

:::: Basic Text Formatting Tags- Lab

file:///D:/APCAS/2021-Even Semester/I MCA/Lab Exercise/Programs/textformatting.xh

Search

Bolded Text

Italized Text

Bolded and Italized Text

<u>Underlined Text</u>

Bigger

Still Bigger Text

Smaller Text

Still Smaller Text

Highlighted Text

In XHTML doctype is ~~not~~ mandatory.

In XHTML elements are ~~not~~ case sensitive.

<u>inserted</u> text

~~Deleted~~ text

This is a ^Superscript^ text

This is a ~Subscript~ text

XHTML stands for **EXtensible HyperText Markup Language**. It is the next step in the evolution of only few differences. This is a cleaner and stricter version of HTML 4.01. If you already know HTML, Consortium (W3C) to help web developers make the transition from HTML to XML. By migrating to and future compatibility of the content.

---

textformatting - Notepad

File Edit Format View Help

```
<h1> Basic Text Formatting Tags </h1>
<b> Bolded Text </b> <br/><br/>
<i> Italized Text </i> <br/> <br/>
<b><i> Bolded and Italized Text </i></b><br/> <br/>
<u> Underlined Text </u><br/> <br/>
<big> Bigger </big>  <br/> <br/>
<big> <big> Still Bigger Text </big> </big>  <br/> <br/>
<small> Smaller Text </small>  <br/> <br/>
<small> <small> Still Smaller Text</small> </small> <br/> <br/>
<mark> Highlighted Text </mark>  <br/> <br/>
In XHTML doctype is <strike> not </strike> mandatory. <br/><br/>
In XHTML elements are <s> not </s> case sensitive. <br/><br/>
<ins> inserted </ins> text <br/> <br/>
<del> Deleted </del> text <br/> <br/> <br/>
This is a <sup> Superscript </sup> text <br/> <br/>
This is a <sub> Subscript </sub> text <br/> <br/>
<p> XHTML stands for <strong> EXtensible HyperText Markup
Language</strong>. It is the next step in the evolution of the
internet.
The XHTML 1.0 is the first document type in the <em>XHTML family
</em>. XHTML is almost identical to HTML 4.01 with only few
differences. This is a cleaner and stricter version of HTML 4.01.
If you already know HTML, then you need to give little attention
to learn this latest version of HTML. XHTML was developed by
```

Ln 24, Col 39   100%   Windows (CRLF)   UTF-8

Type here to search

ENG   02:11 AM   27-02-2021

Monospaced font

- Every browser should have two fonts available: a proportionally-spaced font (such as Times) and a monospaced font (such as Courier).

- Most of the fonts are known as variable-width fonts because different letters are of different widths (for example, the letter 'm' is wider than the letter 'i').

- A monospaced font, also called a fixed-pitch, fixed-width, or non-proportional font, is a font whose letters and characters each occupy the same amount of horizontal space.

- Some examples of text that should be displayed in a monospaced font are computer code, URLs, and commands to enter using the keyboard.

- There are four elements that format text using a monospaced font: **tt, code, kbd, and samp.** These elements represent typewriter text, computer code, keyboard instructions, and sample text respectively.

Other Text Formatting Tags

Some of us had typewriters, and would practice our skills using the phrase `The quick brown fox jumped over the lazy dog`. Now we can do repeated text in Perl: `print "$_\n" x 60;`. Simply type `perl foo.pl` and enter the phrase `sample text`.

*otf - Notepad

File Edit Format View Help

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0
Transitional// EN" "http://www.w3.org/TR/xhtml11-
transitional.dtd">
<HTML xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title>:::: Basic Text Formatting Tags- Lab Exercise- 2
::::</title>
</head>
<body>
<h1> Other Text Formatting Tags </h1>
<p>Some of us had typewriters, and would practice our
skills using the phrase <tt>The quick brown fox jumped
over the lazy dog</tt>. Now we can do repeated text in
Perl: <code>print "$_\n" x 60;</code>. Simply type
<kbd>perl foo.pl</kbd> and enter the phrase <samp>sample
text</samp>.</p>
</body>
</HTML>
```
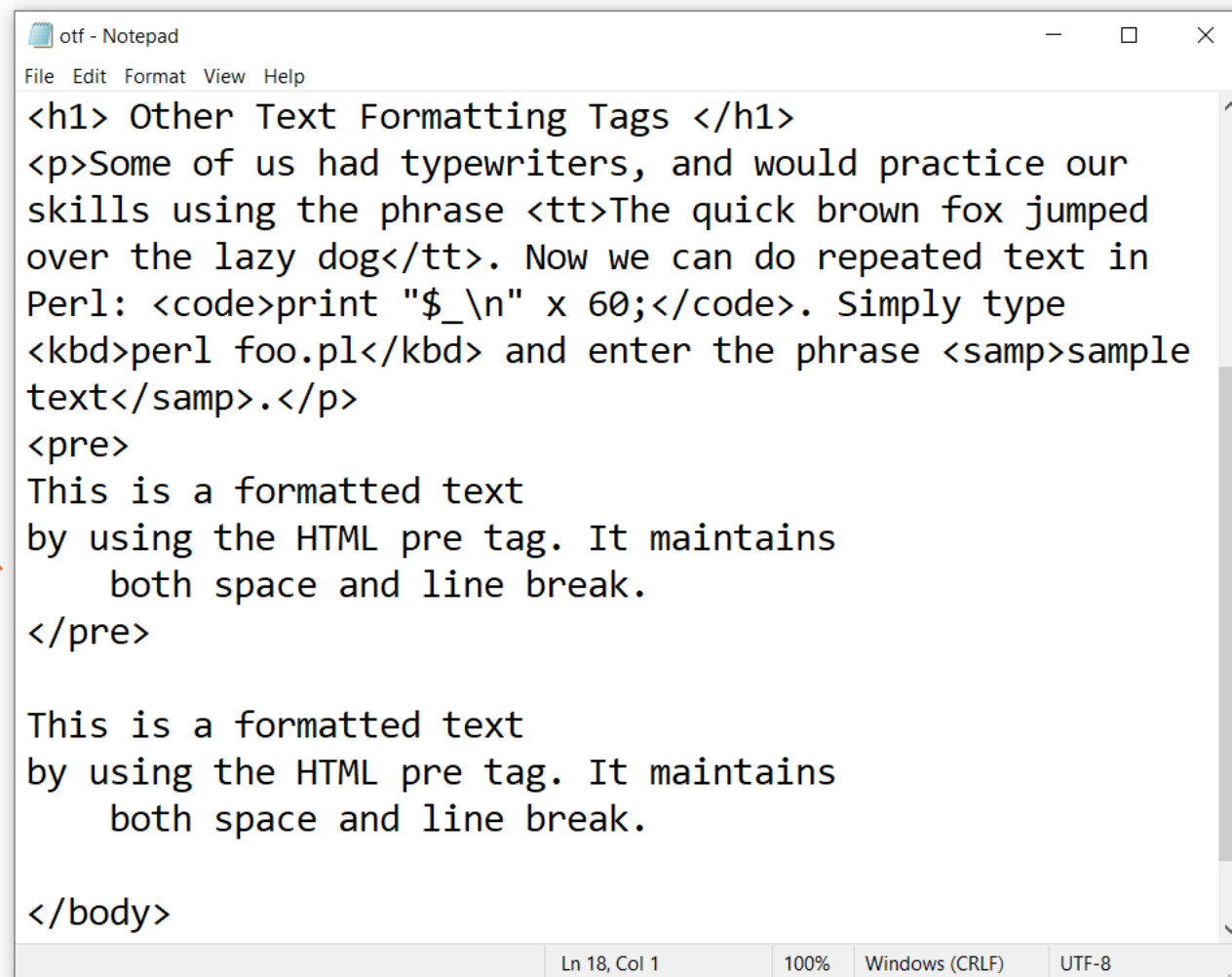
Ln 8, Col 1   100%   Windows (CRLF)   UTF-8

<pre> Tag

- The HTML **<pre>** tag is used for indicating preformatted text.

- The text enclosed within a <pre> tag retains all spacing and returns, and doesn't reflow when the browser is resized. Usually it is displayed in Courier font.

- Simple forms of text formatting, such as bold, italics or links can be used within <pre> tags.

- The HTML <pre> tag also supports the width attribute. The width attribute specifies the desired width of the pre-formatted text.

Other Text Formatting Tags

Some of us had typewriters, and would practice our skills using the phrase `The quick brown fox jumped over the lazy dog`. Now we can do repeated text in Perl: `print "$_\n" x 60;`. Simply type `perl foo.pl` and enter the phrase `sample text`.

```
This is a formatted text
by using the HTML pre tag. It maintains
     both space and line break.
```

This is a formatted text by using the HTML pre tag. It maintains both space and line break.

otf - Notepad

File   Edit   Format   View   Help

```
<h1> Other Text Formatting Tags </h1>
<p>Some of us had typewriters, and would practice our
skills using the phrase <tt>The quick brown fox jumped
over the lazy dog</tt>. Now we can do repeated text in
Perl: <code>print "$_\n" x 60;</code>. Simply type
<kbd>perl foo.pl</kbd> and enter the phrase <samp>sample
text</samp>.</p>
<pre>
This is a formatted text
by using the HTML pre tag. It maintains
     both space and line break.
</pre>


This is a formatted text
by using the HTML pre tag. It maintains
     both space and line break.


</body>
```

Ln 18, Col 1   100%   Windows (CRLF)   UTF-8

<div> and <span>

- The <div> and <span> elements allow you to group together several elements to create sections or subsections of a page.

- The div tag is generally used by web developers to group HTML elements together and apply CSS styles to many elements at once. For example: If you wrap a set of paragraph elements into a div element so you can take the advantage of CSS styles and apply font style to all paragraphs at once instead of coding the same style for each paragraph element.

- <span> tag is used as a generic container of inline elements. It is used for styling purpose to the grouped inline elements (using class and id attribute or inline style).

- The <span> tag does not have any default meaning or rendering.

- The <span> tag can be useful for the following task:
    - To change the language of a part of the text.
    - To change the color, font, background of a part of text using CSS
    - To apply the scripts to the particular part of the text.

:::: Basic Text Formatting Tags- Lab E

file:///D:/APCAS/2021-Even Semester/I MCA/Lab Exercise/Programs/div.xhtml

Search

HOME | **CONTACT** | **ABOUT**

Content Articles

Actual content goes here.....

---

div - Notepad

File Edit Format View Help

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0
Transitional// EN" "http://www.w3.org/TR/xhtml11-
transitional.dtd">
<HTML xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title>:::: Basic Text Formatting Tags- Lab Exercise- 2
::::</title>
</head>
<body>
  <div id = "menu" align = "middle" >
          <a href = "/index.htm">HOME</a> |
          <a href = "/about/contact_us.htm">CONTACT</a> |
          <a href = "/about/index.htm">ABOUT</a>
      </div>
      <div id = "content" align = "left" >
          <h5>Content Articles</h5>
          <p>Actual content goes here.....</p>
      </div>
</body>
</HTML>
```
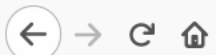
Ln 16, Col 1   100%   Windows (CRLF)   UTF-8

Type here to search

03:37 AM
ENG
27-02-2021

Example of span tag

I have choosen only red, blue, and green colors for my painting.

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0
Transitional// EN" "http://www.w3.org/TR/xhtml11-
transitional.dtd">
<HTML xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title>:::: Basic Text Formatting Tags- Lab Exercise- 2
::::</title>
</head>
<body>
 <h2>Example of span tag</h2>
  <p>I have choosen only
    <span style="color: red;">red</span>,
    <span style="color: blue;">blue</span>, and
    <span style="color: green;">green</span> colors for
my painting.
  </p>
</body>
</HTML>
```

\<address>

- **\<address>** tag is used to specify the authorship information of the article or webpage. It can contain any type of information which is needed such as, URL, physical address, phone number, email, other links, etc.

- The address tag is written as \<address>.

- It usually defines the start of an address.

- All addresses, emails or signatures should be contained in the address tag.

- The address usually appears in italics.

:::: Basic Text Formatting Tags- Lab E   ✕   +

file:///D:/APCAS/2021-Even Semester/I MCA/Examples/address.xhtml

Example of Address tag

Department of Computer Science, Adhiparasakthi College of Arts and Science, Kalavai- 632506.

```
Department of Computer Science,
Adhiparasakthi College of Arts and Science,
Kalavai- 632506.
```

---

address - Notepad

File Edit Format View Help

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0
Transitional// EN" "http://www.w3.org/TR/xhtml11-
transitional.dtd">
<HTML xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title>::::: Basic Text Formatting Tags- Lab Exercise- 2
:::</title>
</head>
<body>
<h2>Example of Address tag</h2>
<address>
Department of Computer Science,
Adhiparasakthi College of Arts and Science,
Kalavai- 632506.
</address>
<pre>
<address>
Department of Computer Science,
Adhiparasakthi College of Arts and Science,
Kalavai- 632506.
</address>
</pre>
</body>
</HTML>
```

Ln 20, Col 1   100%   Windows (CRLF)   UTF-8

Type here to search

03:57 AM
27-02-2021

<abbr>

- HTML <abbr> tag is used to represent an acronym or abbreviation of a longer word or phrase, such as www, HTML, HTTP, etc.

- The content written between <abbr> tags renders with dotted underline in some browser.

- This tag can be used with "title" attribute (optional), and the value of title attribute will be pop-up when the mouse hovers over the content written between <abbr> tag.

Example of Abbrevation tag

HTML XHTML WWW

World Wide Web

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0
Transitional// EN" "http://www.w3.org/TR/xhtml11-
transitional.dtd">
<HTML xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title>::::: Basic Text Formatting Tags- Lab Exercise- 2
:::::</title>
</head>
<body>
<h2>Example of Abbrevation tag</h2>
<abbr title="HyperText Markup language">HTML</abbr>
<abbr title="Extensible HyperText Markup
language">XHTML</abbr>
<abbr title="World Wide Web">WWW</abbr>
</body>
</HTML>
```

\<dfn>

- The HTML Definition element (**\<dfn>**) is used to mark a term that is being defined in the document.

- The definition for the term must be found within the parent of the \<dfn> tag.

- Browsers traditionally render the text found within the \<dfn> tag as italicized text. This tag is also commonly referred to as the \<dfn> element.

HTML is the standard markup language for creating web pages.

HTML is the standard markup language for creating web pages.

HTML is the standard markup language for creating web pages.

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0 Transitional// EN" "http://www.w3.org/TR/xhtml11-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" lang="en">
    <head>
        <title>::::: HTML Definition tag :::::</title>
    </head>
    <body>
<p><dfn>HTML</dfn> is the standard markup language for creating web pages.</p>
<p><dfn title="HyperText Markup Language">HTML</dfn> is the standard markup language for creating web pages.</p>
<p><dfn><abbr title="HyperText Markup Language">HTML</abbr></dfn> is the standard markup language for creating
web pages.</p>
    </body>
</html>
```

<acronym> tag

- <acronym> tag is used with title attribute to contain a full explanation of an acronym content. When you hover the mouse on content, then it will show the explanation of word.

- Although <acronym> tag is not supported by HTML5 but instead of that we can use <abbr> tag and we will get the same result using both tag.

- Though the purpose of this tag is purely for the convenience of the author, its default styling varies from one browser to another:

  - Some browsers, like Internet Explorer, do not style it differently than a <span> element.

  - Opera, Firefox, Chrome, and some others add a dotted underline to the content of the element.

  - A few browsers not only add a dotted underline, but also put it in small caps; to avoid this styling, adding something like font-variant: none in the CSS takes care of this case.
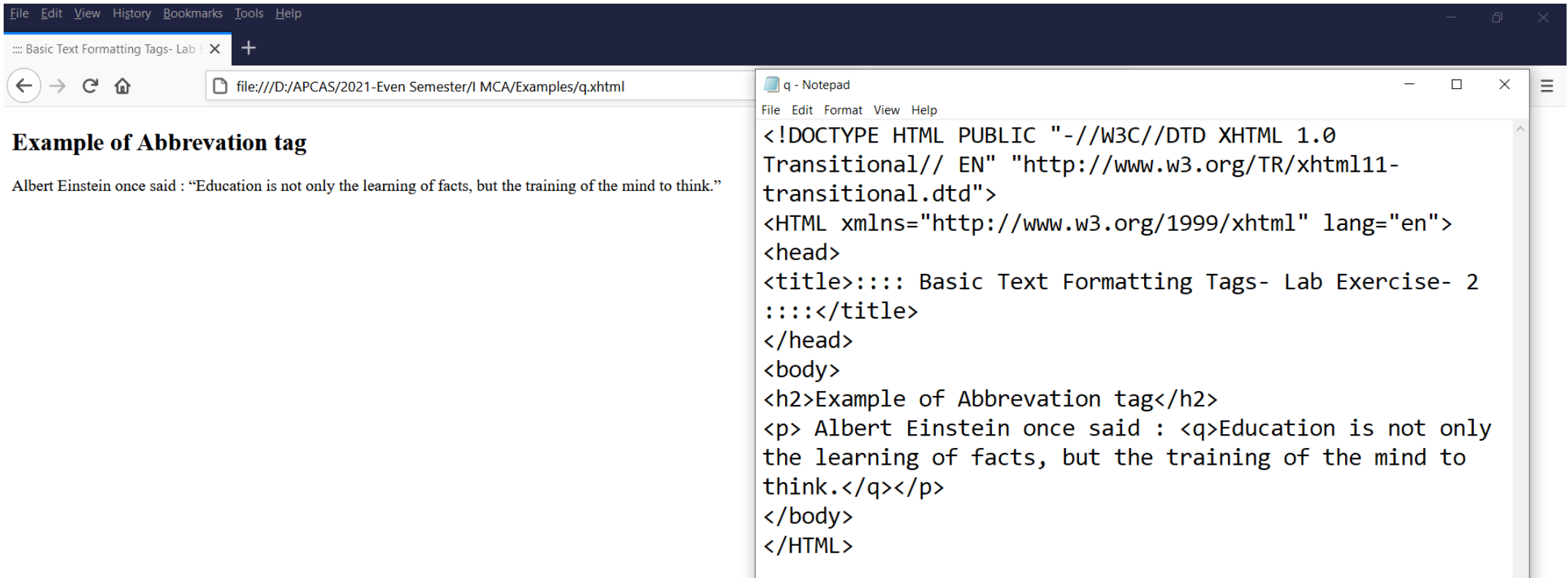
Browser rendering (left window):

Example of Abbrevation tag

Acronym and Abbreviation Example

HTML
HTML XHTML WWW

Notepad source (right window):

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0 Transitional// EN"
"http://www.w3.org/TR/xhtml11-transitional.dtd">
<HTML xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title>::::: Basic Text Formatting Tags- Lab Exercise- 2
:::::</title>
</head>
<body>
<h2>Example of Abbrevation tag</h2>
<h2>Acronym and Abbreviation Example</h2>
    <acronym title="HyperText Markup language">HTML</acronym>
<br/>
<abbr title="HyperText Markup language">HTML</abbr>
<abbr title="Extensible HyperText Markup language">XHTML</abbr>
<abbr title="World Wide Web">WWW</abbr>
</body>
</HTML>
```

HTML Quotes

- The HTML <q> tag is used for indicating short quotations (i.e. quotations that span multiple lines).

<blockquote>

- The XHTML **<blockquote>** Element (or HTML Block Quotation Element) indicates that the enclosed text is an extended quotation (i.e. quotations that span multiple lines). Usually, this is rendered visually by indentation.

- It should contain only block-level elements within it, and not just plain text.

- A URL for the source of the quotation may be given using the cite attribute, while a text representation of the source can be given using the <cite> element.

- The XHTML cite attribute is used to specify the URL of the document that explains the quotes, message or text which describes why the text was inserted.

file:///D:/APCAS/2021-Even Semester/I MCA/Examples/blockquote.xhtml

Search

The HTML blockquote tag is used for indicating long quotations (i.e. quotations that span multiple lines). It should contain only block-level elements within it, and not just plain text. Browsers generally render blockquote text as indented text. If your quoted text needs to display within a non-quoted paragraph, you should use the HTML q tag. Most browsers surround q text with quotation marks.

The HTML blockquote tag is used for indicating long quotations (i.e. quotations that span multiple lines). It should contain only block-level elements within it, and not just plain text. Browsers generally render blockquote text as indented text. If your quoted text needs to display within a non-quoted paragraph, you should use the HTML q tag. Most browsers surround q text with quotation marks.

> "The HTML blockquote tag is used for indicating long quotations (i.e. quotations that span multiple lines). It should contain only block-level elements within it, and not just plain text. Browsers generally render blockquote text as indented text. If your quoted text needs to display within a non-quoted paragraph, you should use the HTML q tag. Most browsers surround q text with quotation marks.

---

blockquote - Notepad

File Edit Format View Help

```
<blockquote cite="http://www.worldwildlife.org/who/index.html">The HTML blockquote tag is used for indicating
long quotations (i.e. quotations that span multiple lines). It should contain only block-level elements within
it, and not just plain text. Browsers generally render blockquote text as indented text. If your quoted text
needs to display within a non-quoted paragraph, you should use the HTML q tag. Most browsers surround q text with
quotation marks.</blockquote>

The HTML blockquote tag is used for indicating long quotations (i.e. quotations that span multiple lines). It
should contain only block-level elements within it, and not just plain text. Browsers generally render blockquote
text as indented text. If your quoted text needs to display within a non-quoted paragraph, you should use the
HTML q tag. Most browsers surround q text with quotation marks.

    <blockquote cite="http://www.worldwildlife.org/who/index.html"><p>The HTML blockquote tag is used for
indicating long quotations (i.e. quotations that span multiple lines). It should contain only block-level
elements within it, and not just plain text. Browsers generally render blockquote text as indented text. If your
quoted text needs to display within a non-quoted paragraph, you should use the HTML q tag. Most browsers surround
q text with quotation marks.</p></blockquote>
```

Ln 21, Col 1   100%   Windows (CRLF)   UTF-8

Type here to search

04:06 AM
01-03-2021

Special Characters

- In HTML, special characters are typically those that can't be easily typed into a keyboard or may cause display issues if typed or pasted into a web page.

- If you plan to use any of the special characters on this page, you should use either the HTML entity name or the HTML entity number. This will ensure that it displays correctly in most/all browsers.

- For example, if you want to display a copyright symbol "©", you should use either © or © in your code.

- All the special characters must be ended with semicolon. (;)

- For Complete references of Special characters refer the following url: https://dev.w3.org/html5/html-author/charref

- Following are some of the special characters.

Character	Entity Number	Entity Name	Description
"	"	"	quotation mark
'	'	'	apostrophe
&	&	&	ampersand
<	<	<	less-than
>	>	>	greater-than
∀	∀	∀	For All
∂	∂	∂	Partial Differential
∃	∃	∃	There Exists
∅	∅	∅	Empty Sets
∇	∇	∇	Nabla
∈	∈	∈	Element Of
∉	∉	∉	Not An Element Of
∋	∋	∋	Contains As Member
∏	∏	∏	N-ary Product
∑	∑	∑	N-ary Summation

Character	Entity Number	Entity Name	Description
Α	Α	Α	Greek Capital Letter Alpha
Β	Β	Β	Greek Capital Letter Beta
Γ	Γ	Γ	Greek Capital Letter Gamma
Δ	Δ	Δ	Greek Capital Letter Delta
Ε	Ε	Ε	Greek Capital Letter Epsilon
Ζ	Ζ	Ζ	Greek Capital Letter Zeta
©	©	©	Copyright Sign
®	®	®	Registered Sign
€	€	€	Euro Sign
™	™	™	Trademark
←	←	←	Leftwards Arrow
↑	↑	↑	Upwards Arrow
→	→	→	Rightwards Arrow
↓	↓	↓	Downwards Arrow
♠	♠	♠	Black Spade Suit
♣	♣	♣	Black Club Suit
♥	♥	♥	Black Heart Suit
♦	♦	♦	Black Diamond Suit

Character	Entity Number	Entity Name	Description
«	«	«	angle quotation mark (left)
¬	¬	¬	negation
	­	­	soft hyphen
¯	¯	¯	spacing macron
°	°	°	degree
±	±	±	plus-or-minus
²	²	²	superscript 2
³	³	³	superscript 3
´	´	´	spacing acute
µ	µ	µ	micro
¶	¶	¶	paragraph
·	·	·	middle dot
¸	¸	¸	spacing cedilla
¹	¹	¹	superscript 1
º	º	º	masculine ordinal indicator
»	»	»	angle quotation mark (right)

Character	Entity Number	Entity Name	Description
¼	¼	¼	fraction 1/4
½	½	½	fraction 1/2
¾	¾	¾	fraction 3/4
¿	¿	¿	inverted question mark
×	×	×	multiplication
÷	÷	÷	Division
			non-breaking space
¡	¡	¡	inverted exclamation mark
¢	¢	¢	cent
£	£	£	pound
¤	¤	¤	currency
¥	¥	¥	yen
¦	¦	¦	broken vertical bar
§	§	§	section
¨	¨	¨	spacing diaeresis

Browser rendered output:

The Copyright Symbol is : ©
The Registered rank is : ®
The Fractional parts are ½ ¼ ¾
The Mathematical Symbols are ÷ <,>,≤,≥,∀, ∂, ∃, ∅
The Spad, Club and hearts are : ♠,♣, ♥,♦
Direction Symbols are : ←↑→↓

Notepad - Specialchar:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0 Transitional// EN"
"http://www.w3.org/TR/xhtml11-transitional.dtd">
<HTML xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title> Special Characters</title>
<meta charset="ISO-8859-1"/>
</head>
<body>
<h2>
<pre>
The Copyright Symbol is : &#169;
The Registered rank is  : &#174;
The Fractional parts are &#189; &#188; &#190;
The Mathematical Symbols are &#247;
&#60;,&#62;,&#x02264;,&#x02265;,&#8704;, &#8706;, &#8707;, &#8709;
The Spad, Club and hearts are : &#9824;,&#9827;, &#9829;,&#9830;
Direction Symbols are : &#8592;&#8593;&#8594;&#8595;
</pre>
</h2>
</body>
</HTML>
```

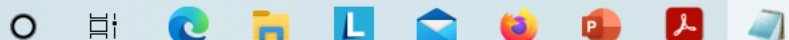# Image maps in XHTML

- The <map> tag is used to define an image map.

- An image map is an image with clickable areas. The areas are defined with one or more <area> tags.

- The idea behind an image map is that you should be able to perform different actions depending on where in the image you click.

- To create an image map you need an image, and some HTML code that describes the clickable areas.

- A clickable area is defined using an <area> element.

- It is mandatory to define the shape of the clickable area, the possible values are:
  - rect - defines a rectangular region
  - circle - defines a circular region
  - poly - defines a polygonal region
  - default - defines the entire region

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0 Transitional// EN"
"http://www.w3.org/TR/xhtml11-transitional.dtd">
<HTML xmlns="http://www.w3.org/1999/xhtml" lang="en">
<head>
<title> :::: Image Map ::::</title>
</head>
<body>
<img src="workplace.jpg" alt="Workplace" usemap="#workmap"/>
<map name="workmap">
  <area shape="rect" coords="34,44,270,350" alt="Computer"
href="https://en.wikipedia.org/wiki/Computer"> </area>
  <area shape="rect" coords="290,172,333,250" alt="Phone"
href="https://en.wikipedia.org/wiki/Telephone"></area>
  <area shape="circle" coords="337,300,44" alt="Coffee"
href="https://en.wikipedia.org/wiki/Coffee"></area>
</map>
</body>
</HTML>
```