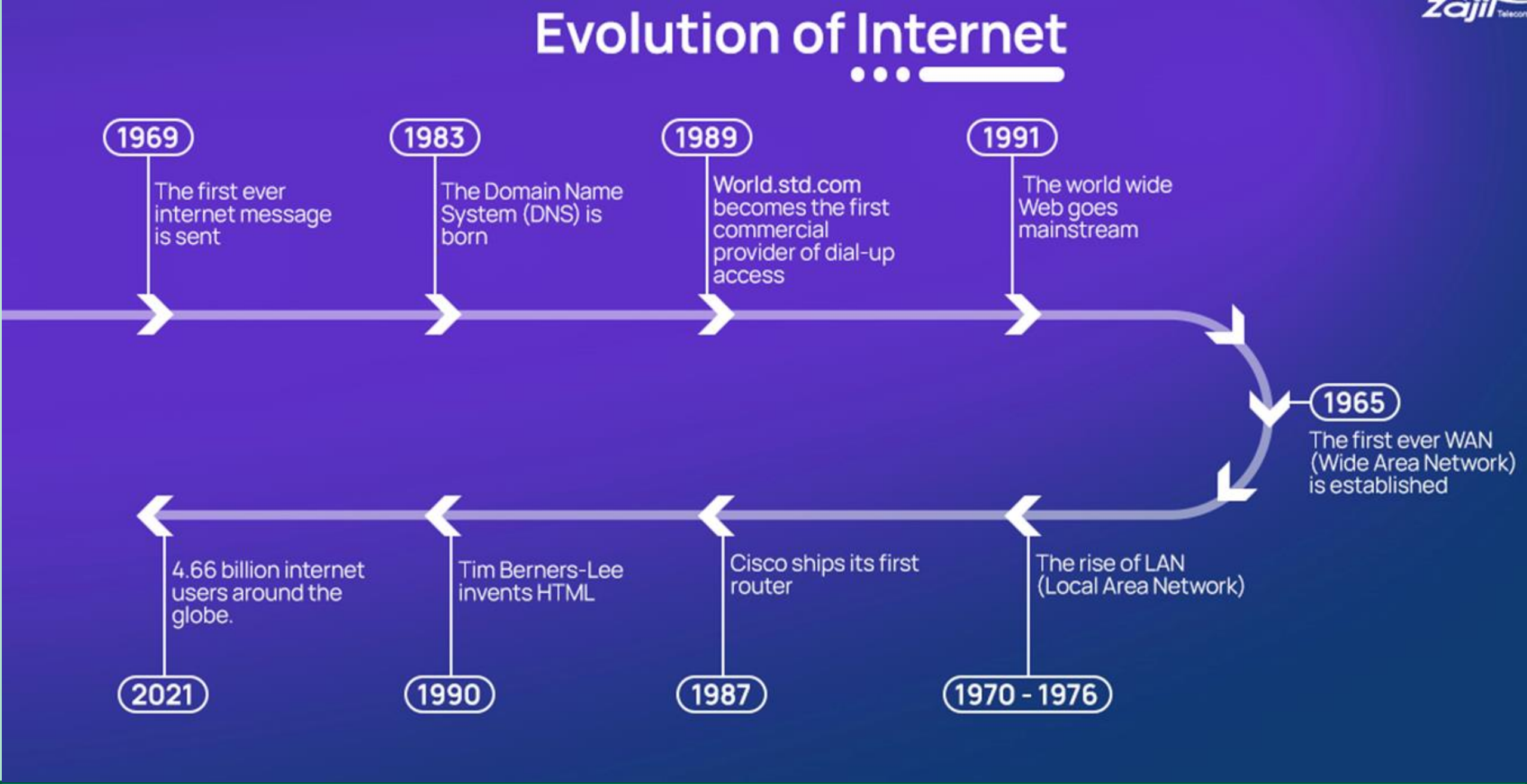
Introduction to the Internet and Web

Evolution of Internet



Internet

- It is the largest network in the world that connects hundreds of thousands of individual networks all over the world.
- The popular term for the Internet is the "**information highway**".
- Rather than moving through geographical space, it moves your ideas and information through cyberspace – the space of electronic movement of ideas and information.

Internet

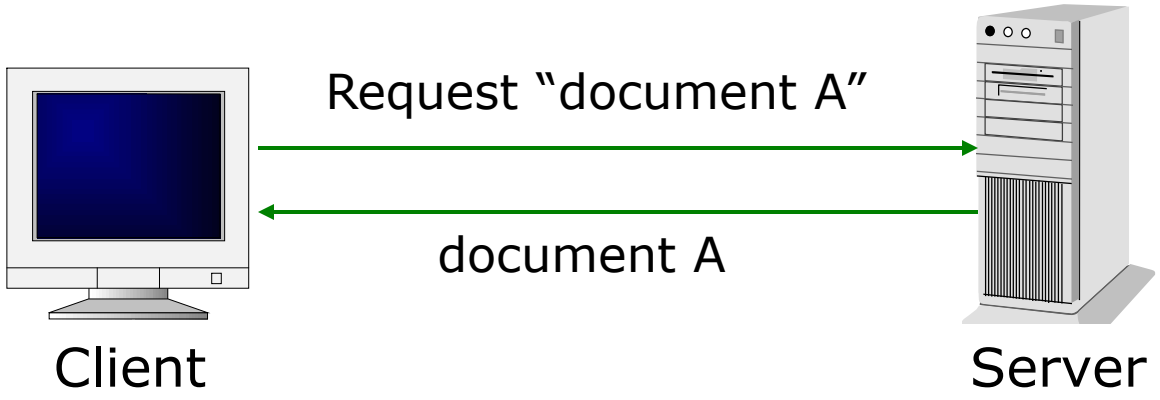
- No one owns it
- It has no formal management organization.
- As it was originally developed by the Department of defense, this lack of centralization made it less vulnerable to wartime or terrorist attacks.
- To access the Internet, an existing network need to pay a small registration fee and agree to certain standards based on the TCP/IP (**Transmission Control Protocol/Internet Protocol**) .

The uses of the Internet

- Send e-mail messages.
- Send (upload) or receive (down load) files between computers.
- Participate in discussion groups, such as mailing lists and newsgroups.
- Surfing the web.

Web Essentials

- **Client**: **web browsers**, used to surf the Web
- **Server** systems: used to supply information to these browsers
- Computer **networks**: used to support the browser-server communication



Internet vs. WWW

- Most people use the two terms interchangeably but they are in fact different.
- The Internet is a vast, international network, made up of computers and the physical connections (wires, routers, etc.) allowing them to communicate.
- **The World Wide Web (WWW or just the Web)** is a collection of software that spans the Internet and enables the interlinking of documents and resources.
  - Provides a way of accessing information on the Internet.

Internet v.s. Web

- **The Internet**: an inter-connected computer networks, linked by wires, cables, wireless connections, etc.
- **Web**: a collection of interconnected documents and other resources.
- The world wide web (**WWW**) is accessible via the Internet, as are many other services including **email, file sharing**, etc.

Web Servers and Clients

- A Web server is a computer that is programmed to send files to browsers on other computers connected to the Internet.
- The Web browser, such as Firefox or Internet Explorer, is the client that sends a request for a Web page.
- The Web server answers the **request and delivers** the requested page to the browser so you can view it.

The World Wide Web (WWW)

- **WWW** is a system of interlinked, hypertext documents that runs over the Internet
- Two types of software:
  - **Client**: a system that wishes to access the information provided by servers must run client software (e.g., web browser)
  - Server: an internet-connected computer that wishes to provide information to others must run server software
  - Client and server applications communicate over the Internet by following a protocol built on top of TCP/IP – HyperText Transport Protocol (**HTTP**)

Basics of the WWW

- **Hypertext**: a format of information which allows one to move from one part of a document to another or from one document to another through **hyperlinks**
- Uniform Resource Locator (**URL**): unique identifiers used to locate a particular resource on the network
- **Markup language**: defines the structure and content of hypertext documents

Web Client: Browser

- Makes HTTP requests on behalf of the user
  - Reformat the URL entered as a valid HTTP request
  - Use DNS to convert server's host name to appropriate IP address
  - Establish a TCP connection using the IP address
  - Send HTTP request over the connection and wait for server's response
  - Display the document contained in the response
    - If the document is not a plain-text document but instead is written in HTML, this involves rendering the document (positioning text, graphics, creating table borders, using appropriate fonts, etc.)

What is Web?

- The **Web (World Wide Web)** consists of information organized into Web pages containing text and graphic images.
- It contains hypertext links, or highlighted keywords and images that lead to related information.
- A collection of linked Web pages that has a common theme or focus is called a **Web site**.
- The main page that all of the pages on a particular Web site are organized around and link back to is called the site's **home page**.

How to access the Internet?

- Many schools and businesses have direct access to the Internet using special high-speed communication lines and equipment.
- Students and employees can access through the organization's local area networks (LAN) or through their own personal computers.
- Another way to access the Internet is through **Internet Service Provider (ISP).**

How to access the Internet?

- To access the Internet, an existing network need to pay a small registration fee and agree to certain standards based on the **TCP/IP (Transmission Control Protocol/Internet Protocol) reference model**.
- Each organization pays for its own networks and its own telephone bills, but those costs usually exist independent of the internet.
- The regional Internet companies route and forward all traffic, and the cost is still only that of a local telephone call.

Internet Service Provider (ISP)

- A commercial organization with permanent connection to the Internet that sells temporary connections to subscribers.
- Examples:
  - BSNL, Vodafone, Idea, Airtel, etc..

How to access the Web?

- Once you have your Internet connection, then you need special software called a browser to access the Web.
- Web browsers are used to connect you to remote computers, open and transfer files, display text and images.
- Web browsers are specialized programs.
- Examples of Web browser: **Netscape Navigator (Navigator) and Internet Explorer.**

Client/Server Structure of the Web

- Web is a collection of files that reside on computers, called **Web servers**, that are located all over the world and are connected to each other through the Internet.
- When you use your Internet connection to become part of the Web, your computer becomes a Web client in a worldwide client/server network.
- A Web browser is the software that you run on your computer to make it work as a web client.

Hypertext Markup Language (HTML)

- The public files on the web servers are ordinary text files, much like the files used by word-processing software.
- To allow Web browser software to read them, the text must be formatted according to a generally accepted standard.
- The standard used on the web is Hypertext markup language (HTML).

Hypertext Markup Language (HTML)

- HTML uses codes, or tags, to tell the Web browser software how to display the text contained in the document.
- For example, a Web browser reading the following line of text:

  <B> A Review of the Book

  <I>Wind Instruments of the 18th Century

  </I>

  </B>

- recognizes the <B> and </B> tags as instructions to display the entire line of text in bold;
- <I> and </I> tags as instructions to display the text enclosed by those tags in italics.

Addresses on the Web:IP Addressing

- Each computer on the internet does have a unique identification number, called an IP (Internet Protocol) address.
- The IP addressing system currently in use on the Internet uses a four-part number.
- Each part of the address is a number ranging from 0 to 255, and each part is separated from the previous part by period,
- For example, 106.29.242.17

IP Addressing

- The combination of the four IP address parts provides 4.2 billion possible addresses (256 x 256 x 256 x 256).
- This number seemed adequate until 1998.
- Members of various Internet task forces are working to develop an alternate addressing system that will accommodate the projected growth.
- However, all of their working solutions require extensive hardware and software changes throughout the Internet.

Domain Name Addressing

- Most web browsers do not use the IP address t locate Web sites and individual pages.
- They use domain name addressing.
- A domain name is a unique name associated with a specific IP address by a program that runs on an Internet host computer.
- This program, which coordinates the IP addresses and domain names for all computers attached to it, is called **DNS (Domain Name System) software**.
- The host computer that runs this software is called a **domain name server.**

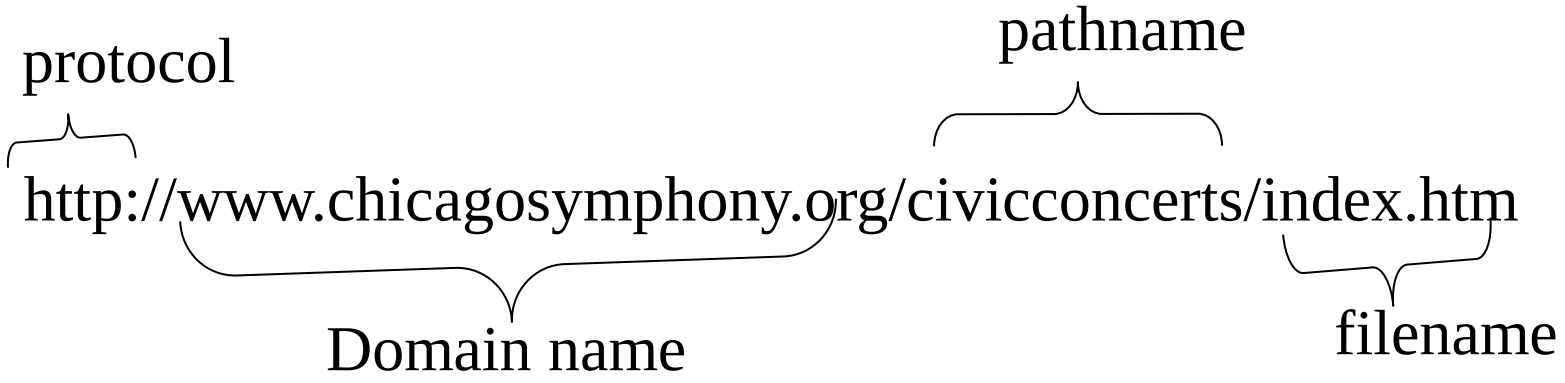
Domain Name Addressing

- Domain names can include any number of parts separated by periods, however most domain names currently in use have only three or four parts.
- Domain names follow hierarchical model that you can follow from top to bottom if you read the name from the right to the left.
- For example, the domain name gsb.uchicago.edu is the computer connected to the Internet at the Graduate School of Business (gsb), which is an academic unit of the University of Chicago (uchicago), which is an educational institution (edu).
- No other computer on the Internet has the same domain name.

Uniform Resource Locators

- The IP address and the domain name each identify a particular computer on the Internet.
- However, they do not indicate where a Web page's HTML document resides on that computer.
- To identify a Web pages exact location, Web browsers rely on Uniform Resource Locator (URL).
- URL is a four-part addressing scheme that tells the Web browser:
  - What transfer protocol to use for transporting the file
  - The domain name of the computer on which the file resides
  - The pathname of the folder or directory on the computer on which the file resides
  - The name of the file

Structure of a Uniform Resource Locators

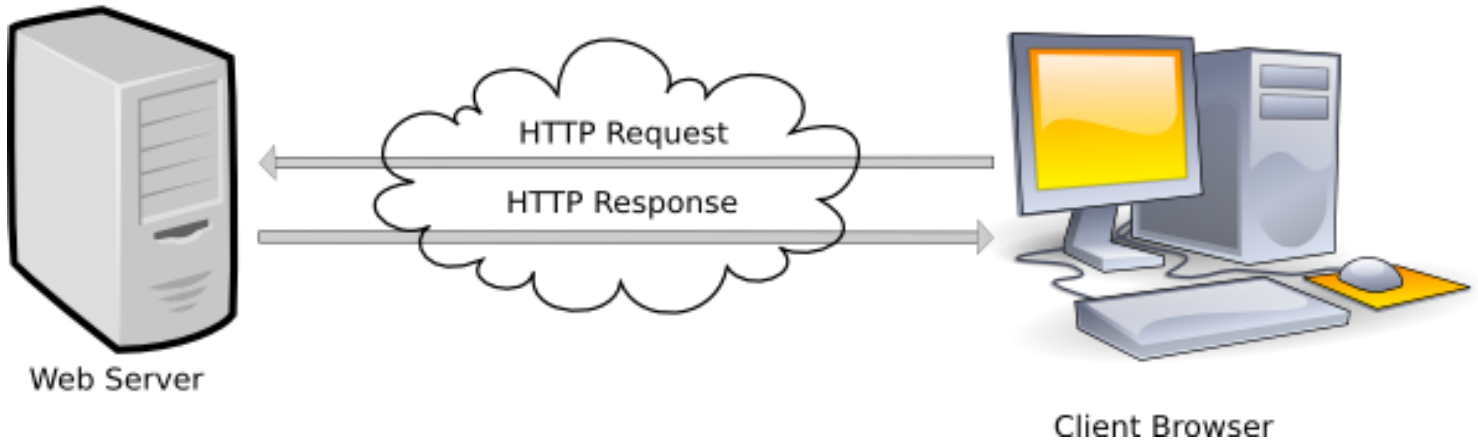


http => Hypertext Transfer Protocol

HTTP

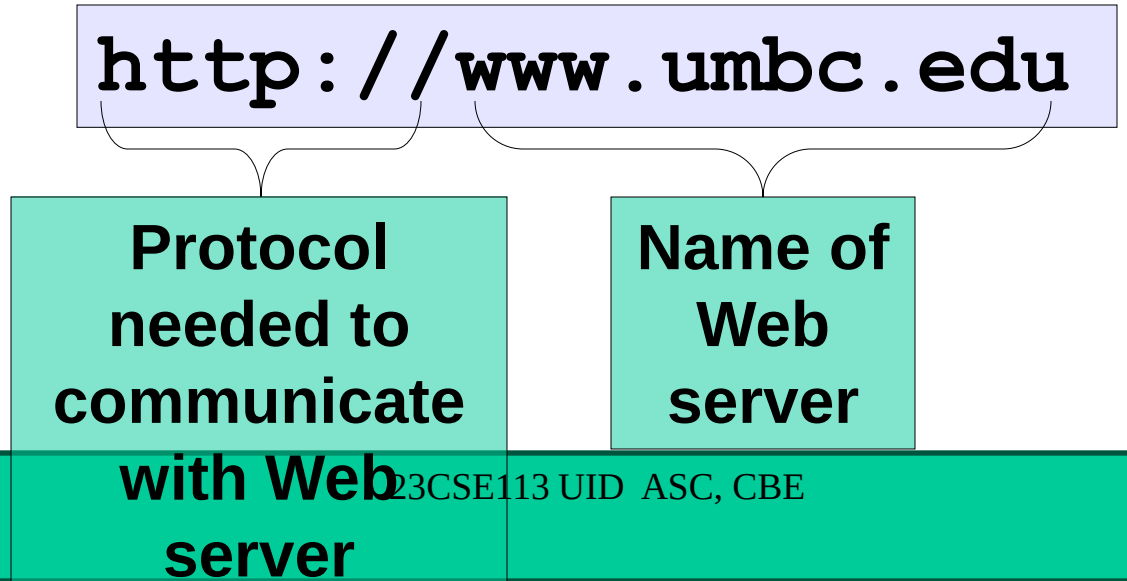
- Stands for HyperText Transfer Protocol
- Allows computers on the WWW to communicate with one another.
- Handles the “request” sent to the Web server and the “response” received from the Web server.

Web Server-Client Diagram



URLs

- Stands for Uniform Resource Locator
- Also called the Web page's address
- You typically type it into your Web browser's location bar when you want to view a Web page



HTML

- Stands for HyperText Markup Language
- Used to create a Web page
- Made up of tags that specify the structure of the document (this section is a heading, this section is a paragraph, etc..)
- An excerpt from a sample HTML document:

```
<html>
  <head>
    <title>Bob's Web page</title>
  </head>
  <body>
```

HTML Tags

- Most HTML tags work in pairs. There is an opening and a closing tag. For example:

```
<p>Some content here.</p>
```

- The ** **...** ** tag displays a paragraph
- ** ** opens the paragraph (opening tag)
- ** ** closes the paragraph (closing tag)
- “Some content here.” will be displayed on the page

Self-closing Tags

- Some HTML tags are self closing. For example:

`<br />`

- The **`<br />`** tag will display a line break.

HTTP

- The transfer protocol is the set of rules that the computers use to move files from one computer to another on the Internet.
- The most common transfer protocol used on the Internet is the Hypertext Transfer Protocol (HTTP).
- Two other protocols that you can use on the Internet are the File Transfer Protocol (FTP) and the Telnet Protocol

How to find information on the Web?

- A number of search tools have been developed and available to you on certain Web sites that provide search services to help you find information.
- Examples:

➢ Yahoo → www.yahoo.com

➢ Excite → www.excite.com

➢ Lycos → www.lycos.com

➢ AltaVista → www/alta-vista.com

➢ MSN WebSearch → www.search.msn.com

How to find information on the Web?

- You can find information by two basic means.
- **Search by Topic** and **Search by keywords**.
- Some search services offer both methods, others only one.
- Yahoo offers both.
  - ➢Search by Topic

    You can navigate through topic lists
  - ➢Search by keywords

    You can navigate by entering a keyword or phase into a search text box.

The Evolution of the Internet

- Internet
  - **Inter**connected **net**work of computer networks
  - ARPAnet
    - Advanced Research Project Agency
    - 1969 – four computers connected
  - NSFnet
    - National Science Foundation
  - Use of the Internet was originally limited to government, research and academic use
  - 1991 Commercial ban lifted

Growth of the Internet

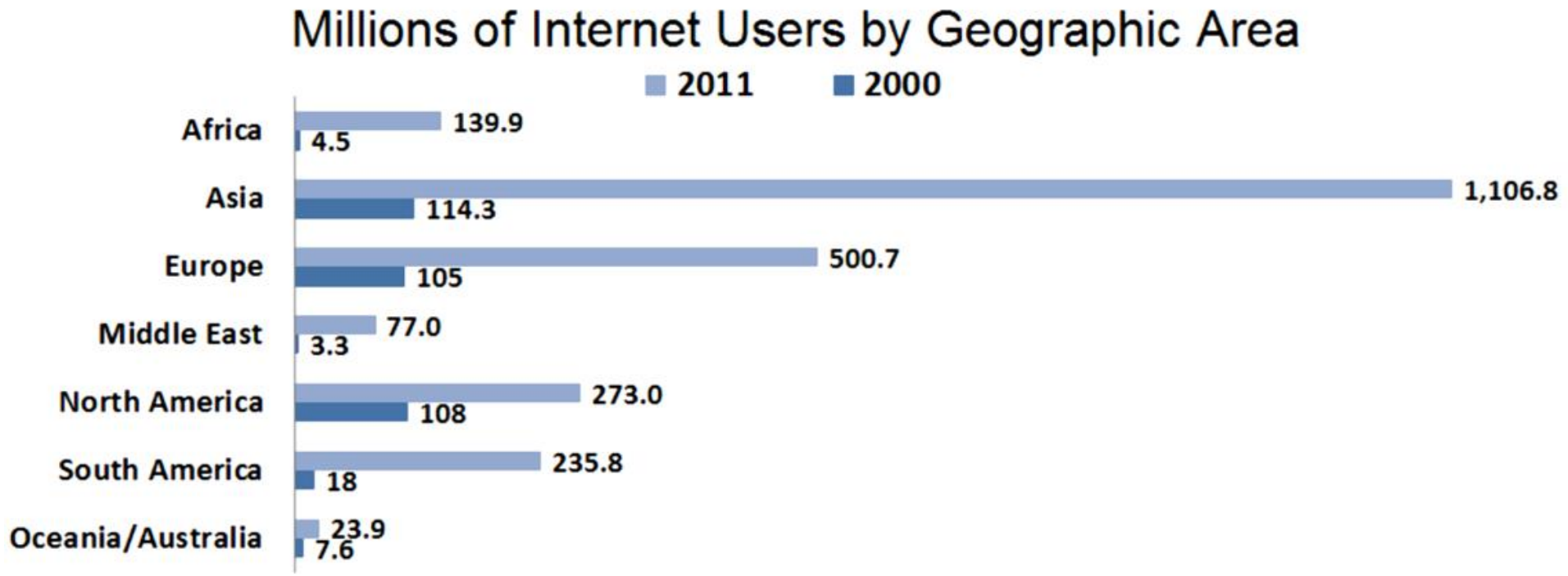


Figure 1.1 Growth of Internet Usage

Statistics from www.internetworldstats.com. Copyright © 2001 - 2012, Miniwatts Marketing Group. All rights reserved worldwide.

Reasons for Internet Growth in the 1990s

- Removal of the ban on commercial activity
- Development of the World Wide Web by Tim Berners-Lee at CERN
- Development of Mosaic, the first graphics-based web browser at NCSA
- Convergence of technologies:
  - Affordable personal computers with GUI Operating Systems
  - Affordable Internet service providers

The World Wide Web

The graphical user interface to information stored on some of the computers connected to the Internet.

Web Standards and the W3C Consortium

W3C – World Wide Web Consortium

- Develops recommendations and prototype technologies related to the Web
- Produces specifications, called Recommendations, in an effort to standardize web technologies
- WAI – Web Accessibility Initiative

Web Accessibility

"The power of the Web is in its universality. Access by everyone regardless of disability is an essential aspect." – Tim Berners-Lee

- Accessible Websites
  provide accommodations that help individuals to individuals with visual, auditory, physical, and neurological disabilities overcome barriers

- WAI – Web Accessibility Initiative
  - Develops accessibility recommendations
  - WCAG 2.0
  - Web Content Accessibility Guidelines
    http://www.w3.org/WAI/WCAG20/quickref/

Web Accessibility

Section 508 of the Rehabilitation Act

- Requires that government agencies must give individuals with disabilities access to information technology that is comparable to the access available to others

Universal Design

"The design of products and environments to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design."

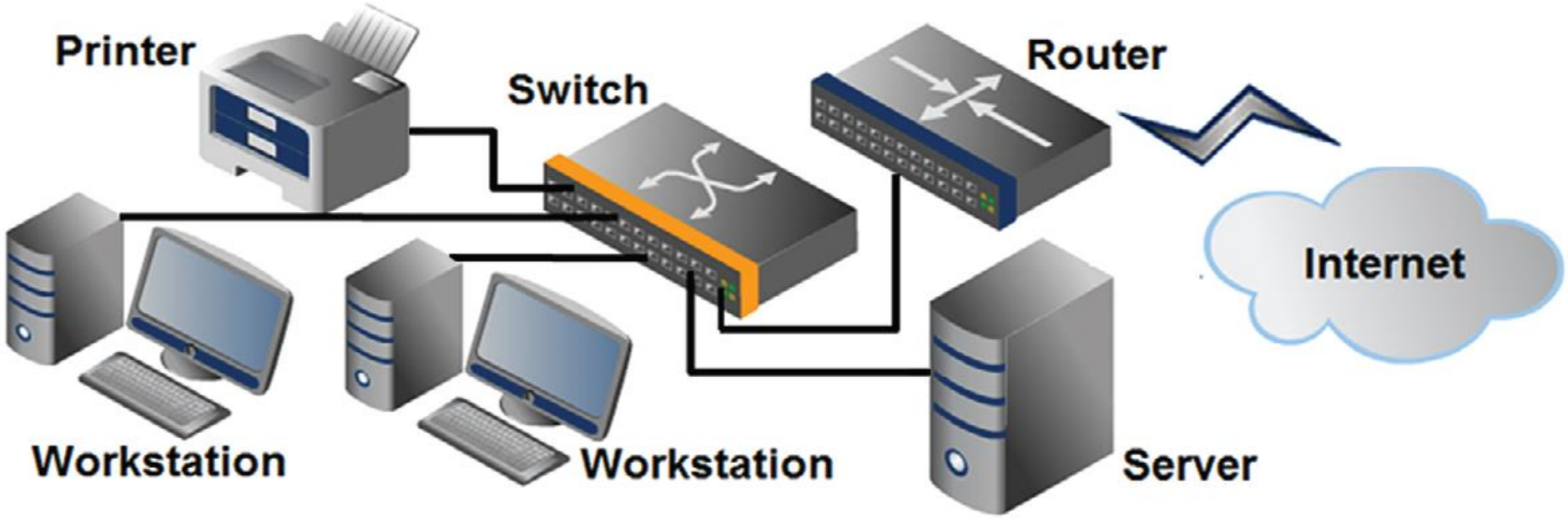
– *The Center for Universal Design*

http://www.design.ncsu.edu/cud/about_ud/about_ud.htm

Information on the Web

- Reliability and information
  - Evaluate the credibility of the site
- Ethical use of information
  - Copyright and the Web

Network Overview



Network

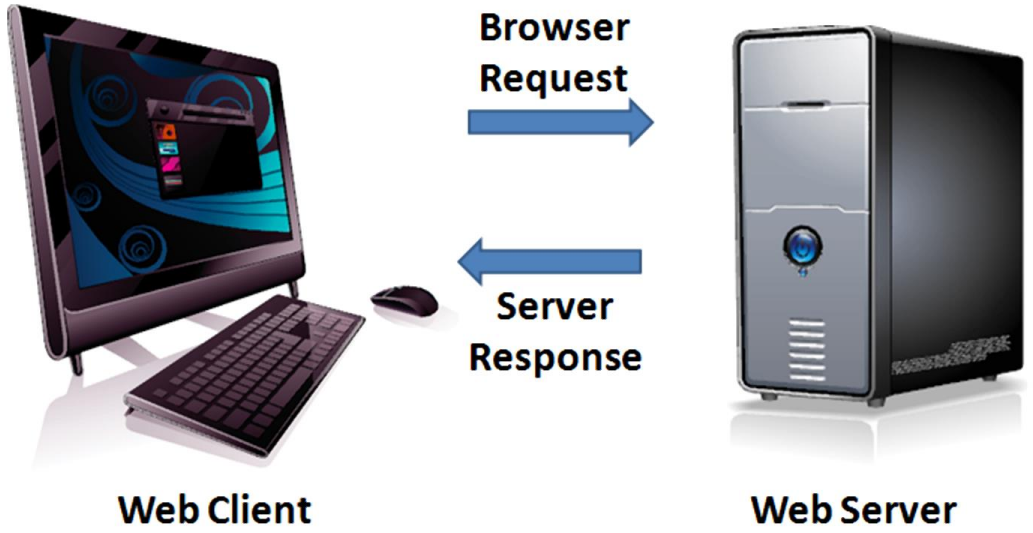
two or more computers connected together for the purpose of communicating and sharing resources

The Client/Server Model

- Client/Server can describe a relationship between two computer programs – the "**client**" and the "**server**".
- Client
  - requests some type of service (such as a file or database access) from the server.
- Server
  - fulfills the request and transmits the results to the client over a network

The Client/Server Model

- The Internet Client/Server Model
  - Client: Web Browser
  - Server: Web Server



Internet Protocols

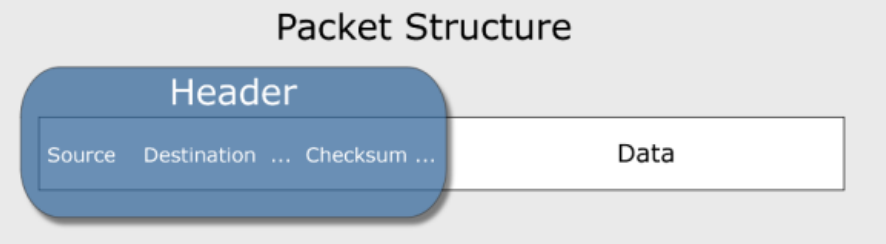
- Protocols
  - Rules that describe the methods used for clients and servers to communicate with each other over a network.

- There is no *single* protocol that makes the Internet and Web work.

- A number of protocols with specific functions are needed.

Common Internet Protocols

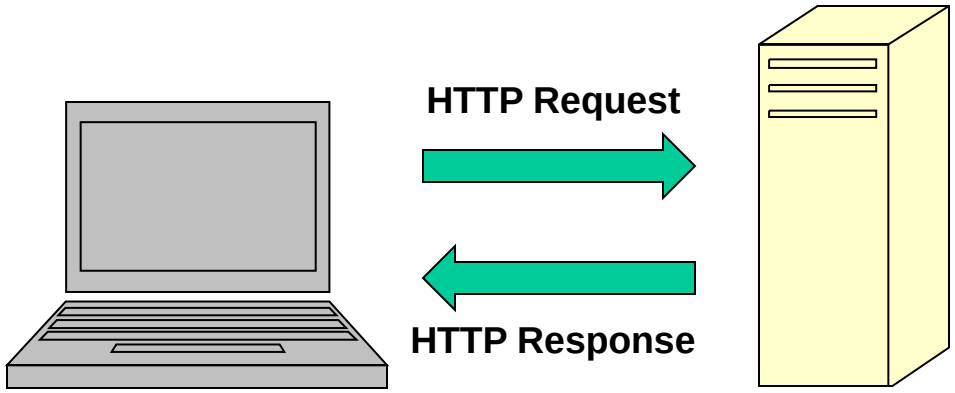
- Official Communication Protocol: TCP/IP



- Specialized Protocols:
  - File Transfer: FTP
  - E-mail: SMTP, POP3, IMAP
  - Websites: HTTP

HTTP
Hypertext Transfer Protocol

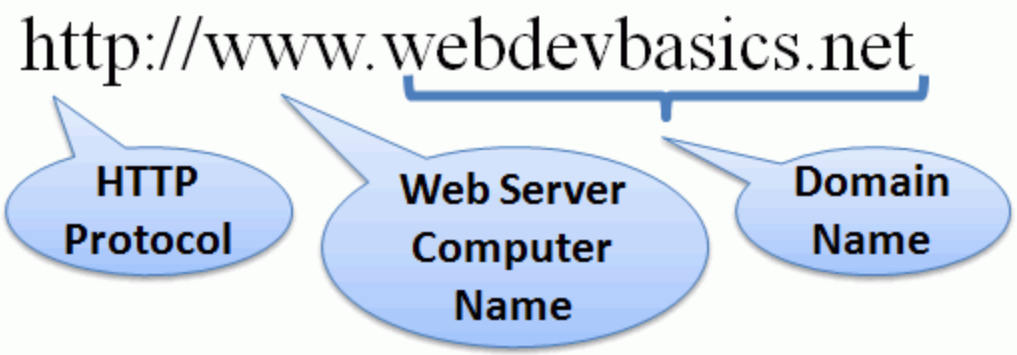
- A set of rules for exchanging files such as text, graphic images, sound, video, and other multimedia files on the Web.

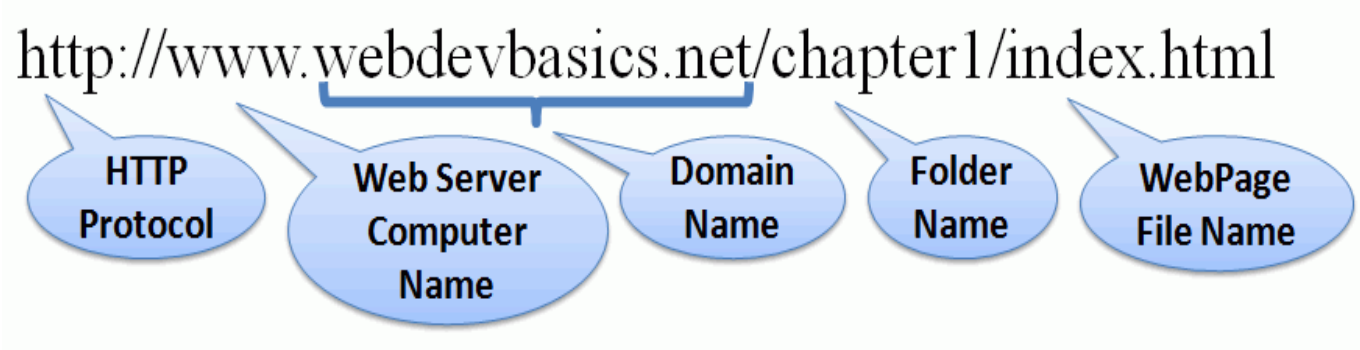


- Web browsers send HTTP requests for web pages and their associated files.

- Web servers send HTTP responses back to the web browsers.

URI
Uniform Resource Indicator





URL

Uniform Resource Locator

Represents the address of a resource on the Internet.

TLD
Top-Level Domain Name

- A top-level domain (TLD) identifies the right-most part of the domain name.

- Some generic TLDs:
  .com, .org, .net, .mil, .gov, .edu, .int, .aero, .asia, .cat, .jobs, .name, .biz, .museum, .info, .coop, .pro, .travel

County Code TLDs

- Two character codes originally intended to indicate the geographical location (country) of the web site.
- In practice, it is fairly easy to obtain a domain name with a country code TLD that is not local to the registrant.
- Examples:
  - .tv, .ws, .au, .jp, .uk
  - See http://www.iana.org/cctld/cctld-whois.htm

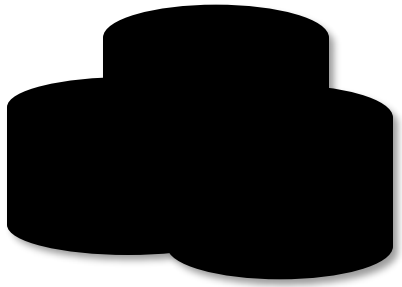
Domain Name System

- The Domain Name System (DNS) associates Domain Names with IP addresses.

Web Browser
requests web page

Domain Name

IP Address



Web
Server

Use TCP/IP
to send HTTP Request

Web Browser
displays web page

Use TCP/IP
to send HTTP Responses
with web page files & images

Markup Languages

- SGML – Standard Generalized Markup Language
  - A standard for specifying a markup language or tag set

- HTML – Hypertext Markup Language
  - The set of markup symbols or codes placed in a file intended for display on a web browser.
    - Element or tag – individual markup code
    - Attribute – modifies the purpose of a tag

Markup Languages (2)

- XML – eXtensible Markup Language
  - A text-based language designed to describe, deliver, and exchange structured information.
  - It is not intended to replace HTML – it is intended to extend the power of HTML by separating data from presentation.

Markup Languages (3)

- XHTML – eXtensible Hypertext Markup Language
  - Developed by the **W3C as the reformulation of HTML 4.0** as an application of XML.
  - It combines the formatting strengths of HTML 4.0 and the data structure and extensibility strengths of XML.

Markup Languages (4)

- HTML 5
  - The next version of HTML 4 and XHTML 1
    - Currently in draft status
    - Incorporates features of both HTML and XHTML
    - Adds new elements
    - Eliminates some elements
    - Intended to be backward compatible
  - http://www.w3.org/html/

Your First HTML5 Web Page: index.html

```
<!DOCTYPE  html">
<html lang="en">
<head>
  <title>Page Title Goes Here</title>
  <meta charset="utf-8">
</head>
<body>
 ... body text and more HTML tags go here ...
</body>
</html>
```

Under the Hood of a Web Page

DTD – describes the markup language syntax(Document Type Definition)

HTML element– contains the web page document

Head element – contains the head section

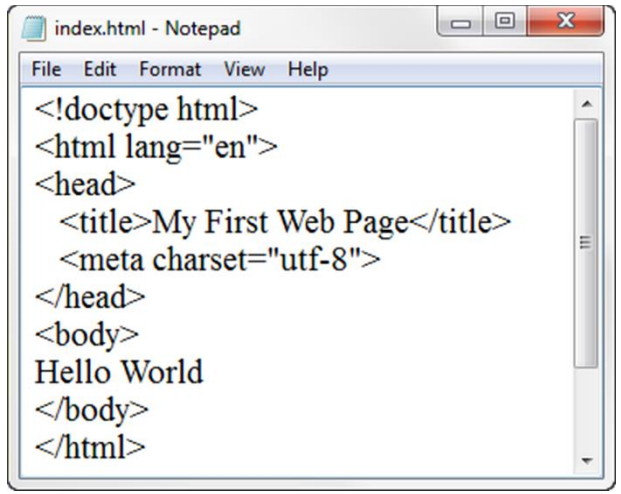
The head section contains information that describes the web page document

Title element– Text displays in title bar of window

Meta element – describes the character encoding

Body element – contains the body section

The body section contains the text and elements that display in the browser viewport.



Static Web: HTML/XHTML, CSS

- **HTML** stands for **H**yper**T**ext **M**arkup **L**anguage
  - It is a text file containing small markup tags that tell the Web browser how to display the page
- **XHTML** stands for e**X**tensible **H**yper**T**ext **M**arkup **L**anguage
  - It is identical to HTML 4.01
  - It is a stricter and cleaner version of HTML
- **CSS** stands for **C**ascading **S**tyle **S**heets
  - It defines how to display HTML elements

Why Programmability?

- What's the drawback to simple document model?
  - Static
  - Assume that documents are created before they are requested
- What are examples of information that might be part of web documents that may not be known before they are requested?

Client-Side Programmability

- Scripting language: a lightweight programming language
- Browser scripting: **JavaScript**
  - Designed to add interactivity to HTML pages
  - Usually embedded into HTML pages
  - What can a JavaScript Do?
    - Put dynamic text into an HTML page
    - React to events
    - Read and write HTML elements
    - Validate data before it is submitted to a server
    - Create cookies
    - …

Server-Side Programmability

- The requests cause the response to be generated
- Server scripting:
  - **CGI/Perl**: Common Gate Way Interface (*.pl, *.cgi)
  - PHP: Open source, strong database support (*.php)
  - ASP: Microsoft product, uses .Net framework (*.asp)
  - Java via JavaServer Pages (*.jsp)
  - …

What's Ahead?

- HTML, XHTML
- CSS
- Simple client-side interactivity (JavaScript)
- Simple server-side interactivity (CGI/Perl) – Later Courses
- We will not "teach" these languages
- We will provide an overview of the basics, and learn how to use the web resources to help build a web page

Summary

This chapter provided a brief overview of Internet, Web, and introductory networking concepts along with your very first web page.