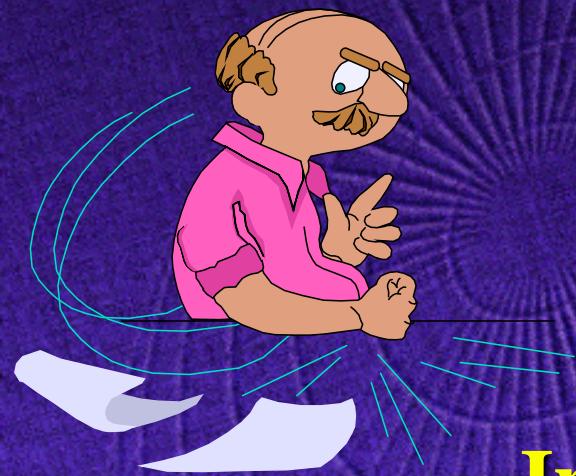


Internet Client-Server Systems

4020 A

Instructor: Jimmy Huang
jhuang@yorku.ca

<http://www.yorku.ca/jhuang/4020A.html>



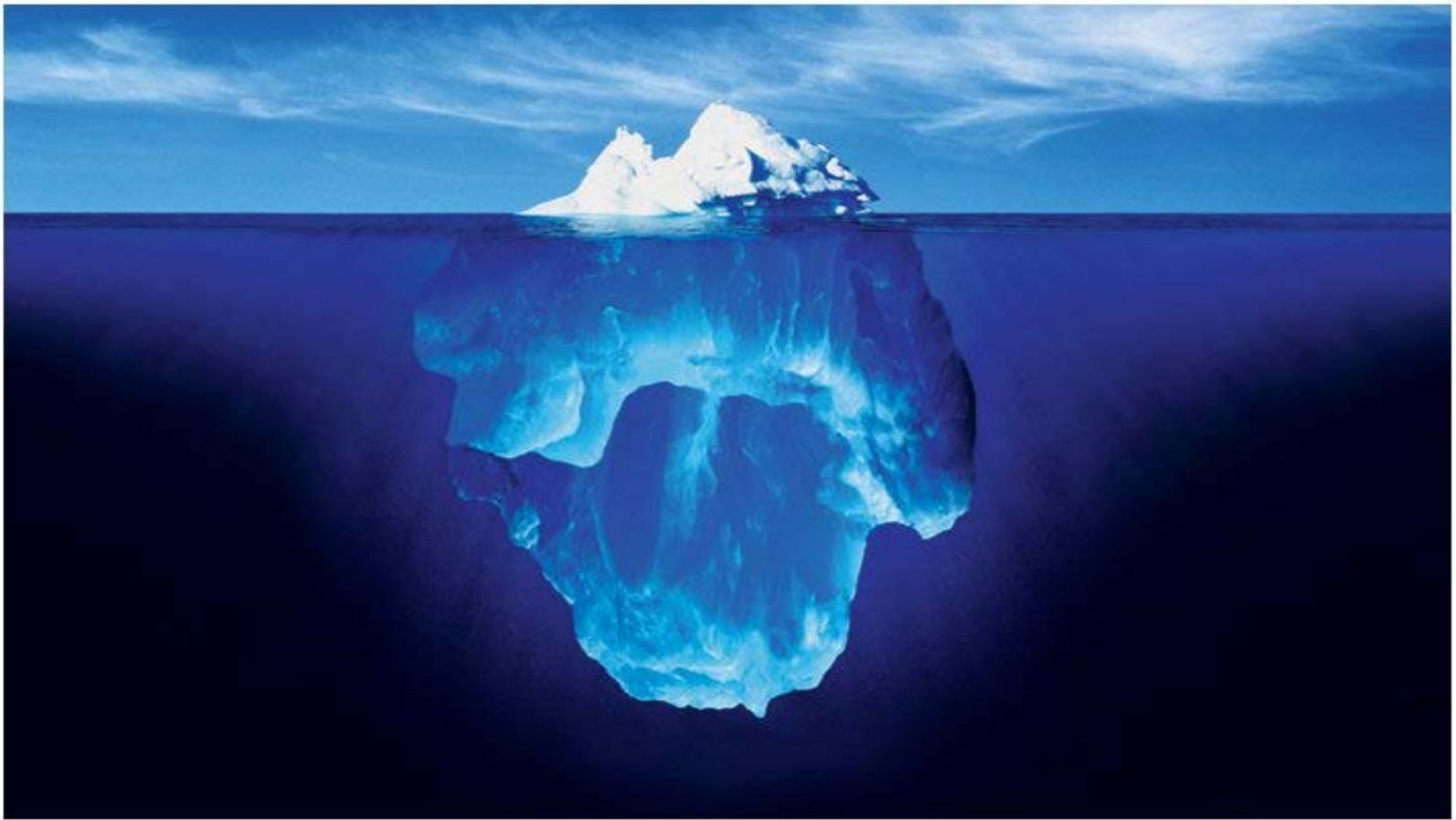
Motivation

- ◆ **Web-based Knowledge & Data Management**
 - A huge amount of Web data
 - how to organize, retrieve them, how discover interesting patterns and how to make a recommend from them
- ⇒ **Web Search Engine**
- ⇒ **Uber Taxi and Didi Chuxing**
- ⇒ **Amazon, Alibaba, Tencent, JD.com**
- ⇒ **Web Blog Analysis**
- ⇒ **Spam Email Detection**
- ⇒ **Online Electronic Medical Data Analysis**
- ⇒ **Electronic Health Care and eHealth**
- ⇒ **Social Network Analysis**

Examples of Web Search Engines



Examples of Web



Some Internet Related Research Projects for Students

- ◆ “Context-Sensitive and Task-Oriented Adaptive Information Retrieval”. This project is supported by NSERC Individual Discovery Grant (2007 - 2019)
- ◆ “Analyzing and Searching Medical Data for Cost Effective Health Care”. This project is supported by Early Researcher Award/Premier’s Research Excellence Award
- ◆ “Finding Best Evidence for Evidence-based Best Practice Recommendations in Health Care”. This project is supported by NSERC Collaborative R&D Grant
- ◆ Other research projects will also be available such as IBM SUR and OCE projects etc.
- ◆ Advanced Data Mining and Machine Learning Technologies for Next Generation eHealth Decision Support Systems



Course Objectives

- ◆ This course will cover both programming aspects of internet applications and advanced topics of Web technology, such as information retrieval, Web search and Web mining.
- ◆ The major objective of this course is for the students to learn how to analyze, design and implement internet applications and obtain a solid grasp of how techniques in Web technology can be applied to solve problems.

Course Content

- ◆ Introduction to client-server systems, WWW and Web technology
- ◆ HTML, XHTML and XML
- ◆ Servlets and Java Server Pages (JSP)
- ◆ JDBC and Java Script
- ◆ E-business and E-commerce
- ◆ Web mining, Web server logs and Web session
- ◆ Information Retrieval (indexing, search and ranking)
- ◆ Web search and Web crawler

Recommended Reading

- ◆ *Internet and World Wide Web: How to Program (2nd edition)*
 - Deitel M., Deitel J., Nieto T. R.
 - Prentice Hall, 2002
- ◆ *Advanced Java 2 Platform: How to Program*
 - Deitel M., Deitel J., Santry S. E.
 - Prentice Hall, 2002
- ◆ *Foundations of Web Technology*
 - Sarukkai R.
 - Kluwer Academic Publishers, 2003
- ◆ *Mining the World Wide Web: An Information Search Approach*
 - Chang G., Healey M. J. et al.
 - Kluwer Academic Publishers, 2003

Evaluation and Grading

- ◆ **3 assignments** (due through the term)
 - Assignment 1: will be announced soon (25%)
 - Assignment 2: TBA (25%)
 - Assignment 3: TBA (20%)
- ◆ **Final Exam** (30%) Dec

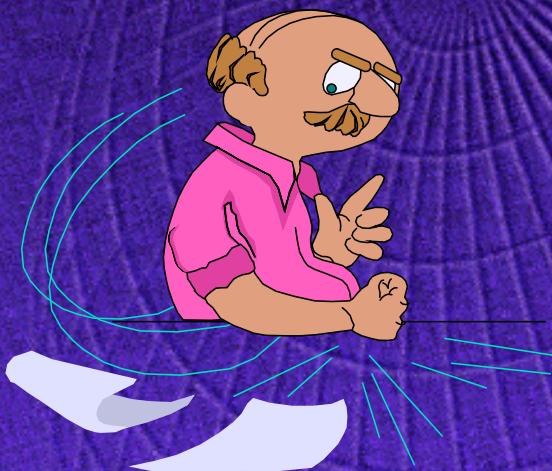
Group Assignments

- ◆ **Each group consists of 4 or 5 members.**
- ◆ **There is a team leader in each group.**
- ◆ **The team leader should send me an email telling me how many people are in the group, their names, student Ids, email addresses and course section ID as soon as possible (before Sept 12).**
- ◆ **Your group assignments will be evaluated in terms of your programs, assignment reports and class presentations.**

Class Time and Office Hours

- ◆ Class time:
 - Thursday: 19:00pm – 22:00pm
- ◆ Class location: Accolade West 008
- ◆ Office hours (**TEL 3048**)
 - By appointment
- ◆ TA name: Sadra Abrishamkar (PhD candidate)
- ◆ TA office hours: to be arranged

Introduction to Client-Server Systems, WWW and Web Technology

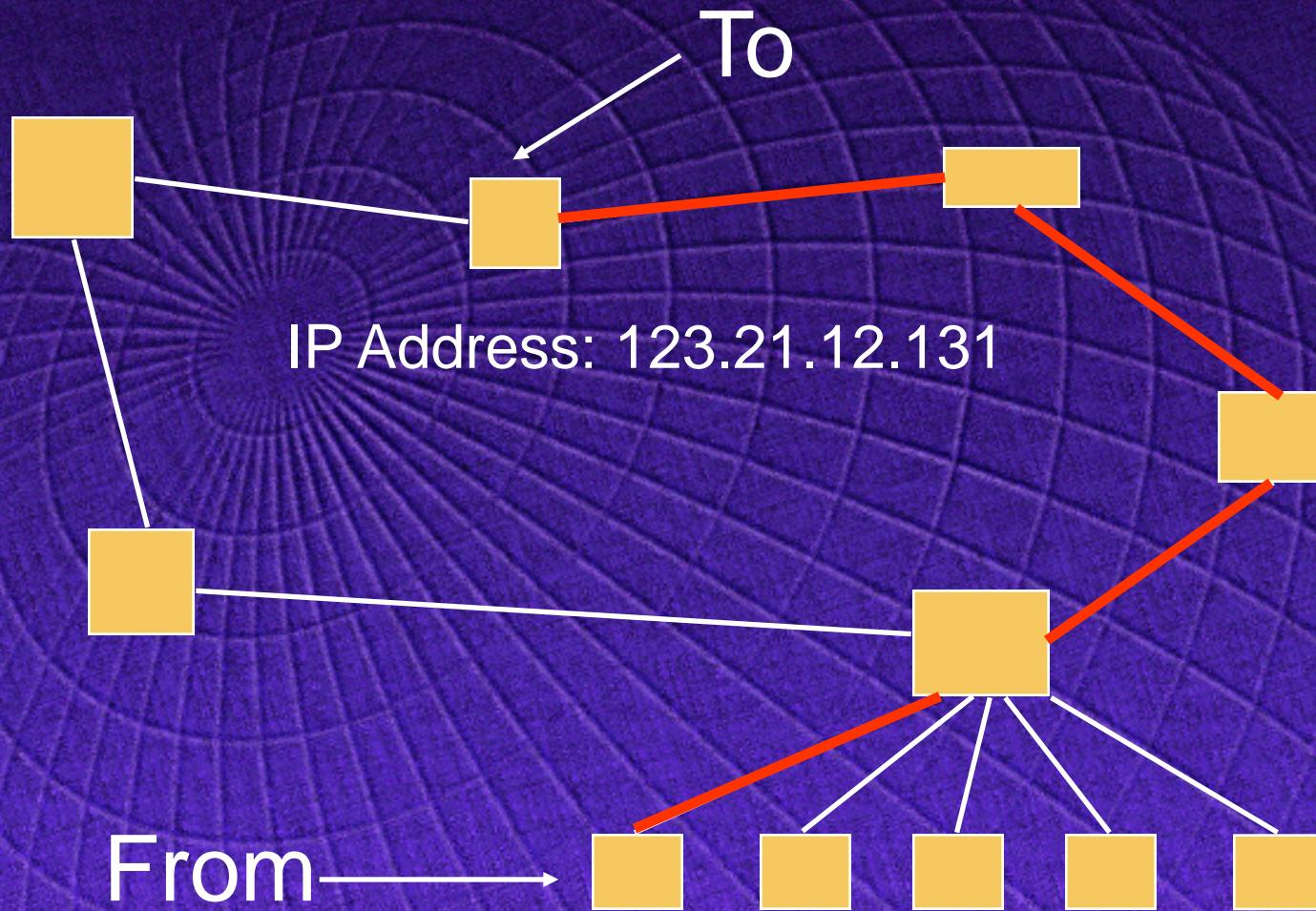


Week 1

Outline

- ◆ The Internet
- ◆ The Web
- ◆ What makes the Web work?
 - HTTP
 - URL
 - HTML
 - CGI
- ◆ Example of a Web page
- ◆ Summary

The Internet



The Internet

Worldwide collection of interconnected networks.

- ◆ Began in late '60s in ARPANET, a US project, investigating how to build networks that could withstand partial outages.
- ◆ Starting with a few nodes, Internet estimated to have over 100 million users in 1997, and over 270 million users in over 100 countries in 1998, with one million new users joining each month.

Historical View: Internet

- ◆ 1969 - Telnet
- ◆ 1970 - 4 computers
 - Stanford, UCLA, UC Santa Barbara, U Utah
- ◆ 1971 - FTP
- ◆ 1983 - 562 computers on the internet
- ◆ 1993 - 1.2 million computers on the internet
- ◆ 1999 -

Outline

- ◆ The Internet
- ◆ **The Web**
- ◆ What makes the Web work?
 - HTTP
 - URL
 - HTML
 - CGI
- ◆ Example of a Web page
- ◆ Summary

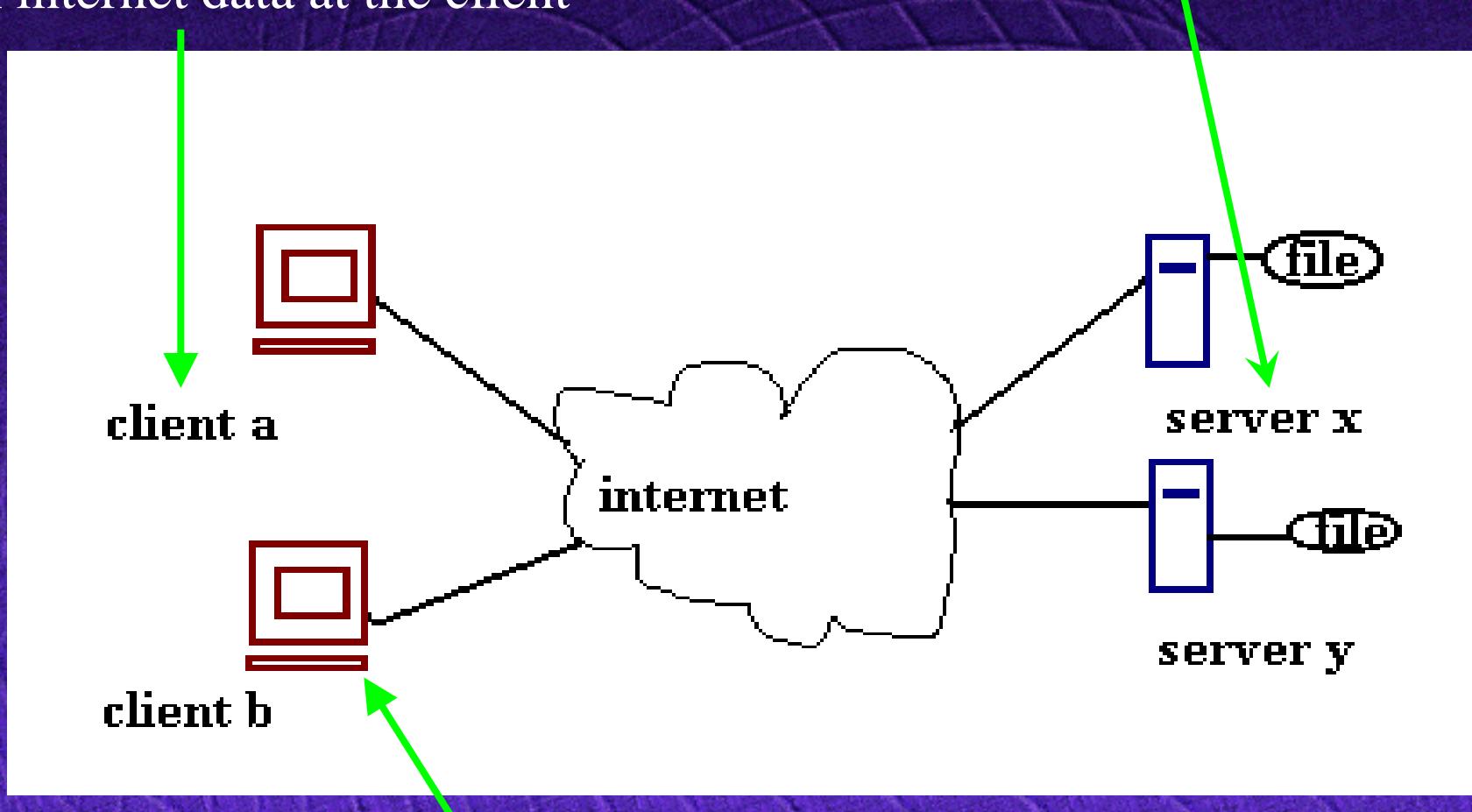
The Web

- ◆ World-Wide Web (Web, WWW)
 - networked information system that provides a simple way of browsing different types (text, pictures, video, audio, etc.) of information on the Internet using hyperlinks.
- ◆ Web pages
 - electronic documents that typically contains several types of information accessible via the World Wide Web
- ◆ Web sites
 - a collection of related Web pages of a certain individual, group, or organization.
- ◆ The Web uses a client/server model

Client-Server Model

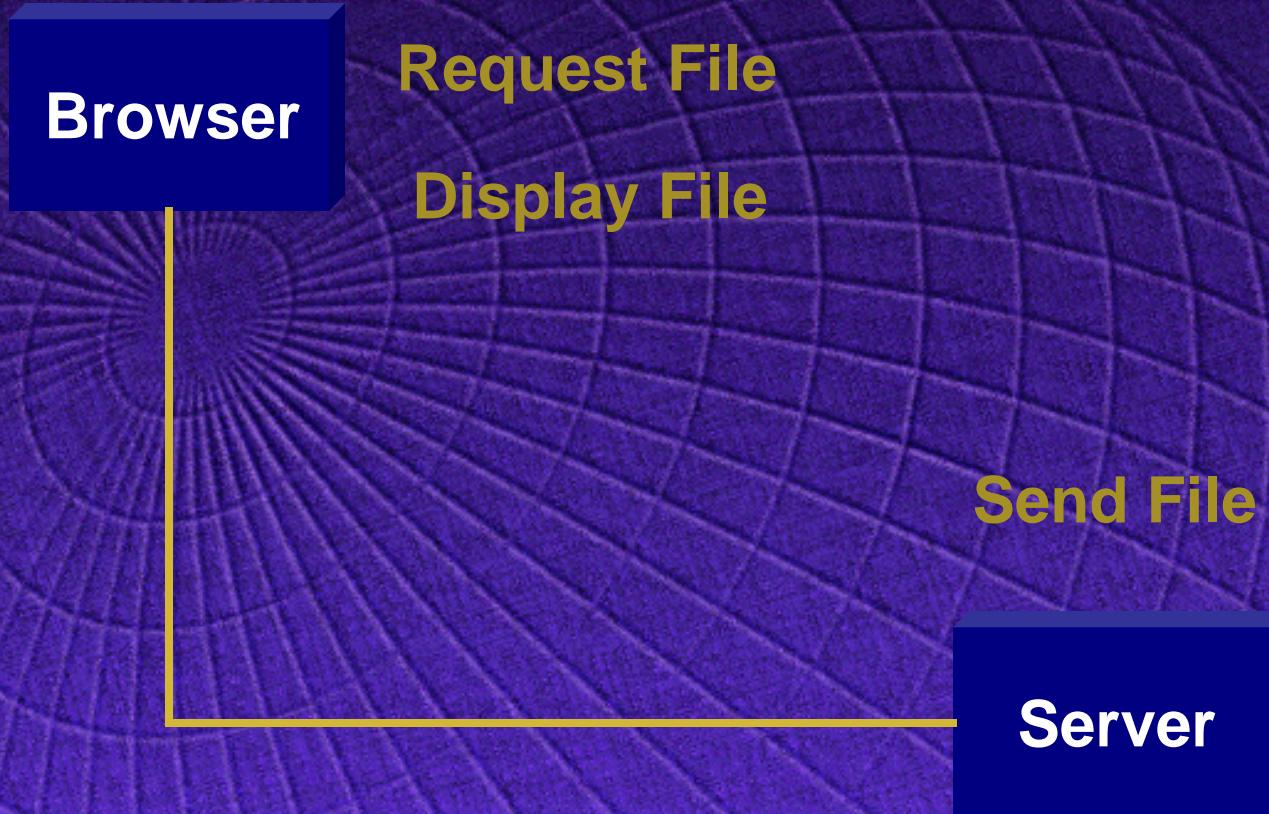
Browser - software to interact with internet data at the client

machine that services internet request



machine that initiates internet request

Client/Server Interaction



What is a Web Server?

Web server

- ◆ computer running application software that listens and responds to a client computer's request made through a web browser
- ◆ machine that hosts web pages and other web documents
- ◆ provides web documents and other online services using HTTP

What is a Web Browser?

Web browser

- ◆ application software that is used to locate and issue a request for the page on the web server that hosts the document
- ◆ It also interpret the page sent back by the web server and display it on the monitor of the client computer
- ◆ computer program that lets you view and explore information on the World Wide Web

Web Browsers



- ◆ Microsoft Internet Explorer – browser integrated with the Windows operating system. Mac versions are available.
- ◆ Netscape Navigator - available for Windows, Mac, and Unix platforms.
- ◆ Opera – one of the alternatives to the two most popular browser mentioned above
- ◆ Mozilla – open source web browser software
- ◆ Lynx - popular Unix text-based browser
- ◆ Google Chrome is a Google browser that combines a minimal design with sophisticated technology.



Outline

- ◆ The Internet
- ◆ The Web
- ◆ **What makes the Web work?**
 - HTTP
 - URL
 - HTML
 - CGI
- ◆ Example of a Web page
- ◆ Summary

What Makes the Web Work?

The Web relies on these mechanisms:

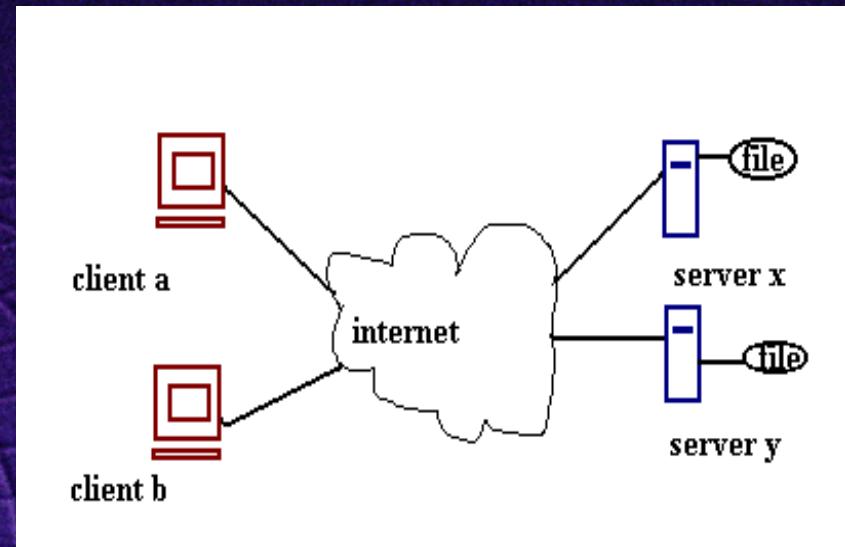
- ◆ Protocols - set of standards used to access resources via the Web
- ◆ Universal Resource Locator (URL) - uniform naming scheme for Internet resources
- ◆ HTML - Document formatting language used to design most Web pages
- ◆ CGI - Common Gateway Interface
- ◆ Servlet - Application run by a server connected to the WWW. It is one of the most popular avenues for Java development today.

Protocols

- ◆ Standard set of rules that governs how computers communicate with each other, i.e. HTTP, FTP and SMTP.
- ◆ HTTP (HyperText Transfer Protocol) is the underlying protocol used to transmit information over the Web.
- ◆ HTTP is based on request-response paradigm:
 - Connection: Client establishes connection with Web server.
 - Request: Client sends request to Web server.
 - Response: Web server sends response (HTML document) to client.
 - Close: Connection closed by Web server.

HTTP Connection

- ◆ 1. Client
 - makes an HTTP request for a web page
 - makes a TCP/IP connection
- ◆ 2. Server accepts request
 - sends page as HTTP
- ◆ 3. Client downloads page
- ◆ 4. Server breaks the connection



Uniform Resource Locators (URLs)

- ◆ Identifies the file to request
 - Specifies server and file
 - Defaults used for missing values



Note: Not all URLs will have the directory and filename



HyperText Markup Language (HTML)

- ◆ Hypertext
 - presents and relates information as hyperlinked documents that point to other documents or resources.
- ◆ HTML
 - A standard markup language that defines a hypertext document.
 - A simple, powerful, platform-independent document language.
 - Specifies what displays should look like
 - Browser interprets HTML
 - Same HTML file often looks different across browsers
 - HTML files are the source files of Web pages

HTML File Structure

```
<HTML>
<HEAD>
<TITLE>Page Title</TITLE>
</HEAD>
<BODY>
Stuff
</BODY>
</HTML>
```

What About Graphics?

- ◆ An HTML file can refer to an image file

Here is a nice picture:

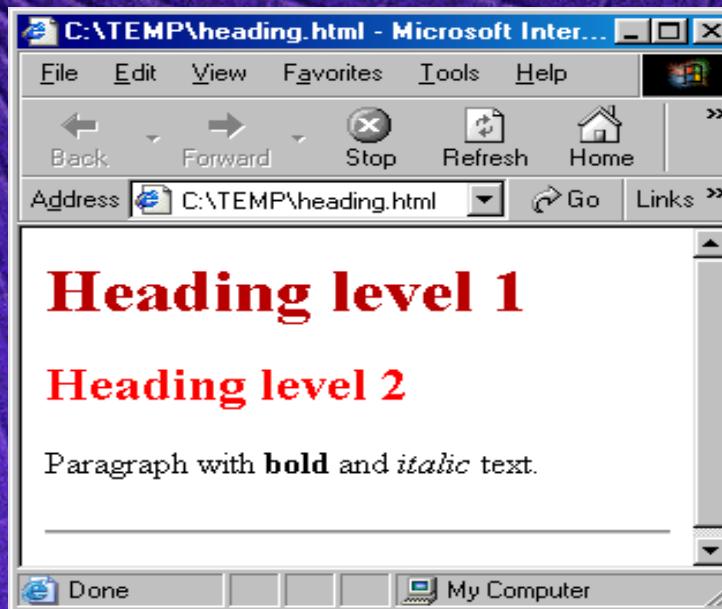
What About Hyperlinks?

- ◆ An HTML file can refer to another HTML file

```
<h2>Teaching</h2>  
  
<p><a href="http://ai.uwaterloo.ca/3421.html">  
COSC 3421 Fall 2002</a></p>  
  
<p><a href="http://ai.uwaterloo.ca/3221.html">  
COSC 3221 Winter 2003</a></p>
```

Simple Formatting

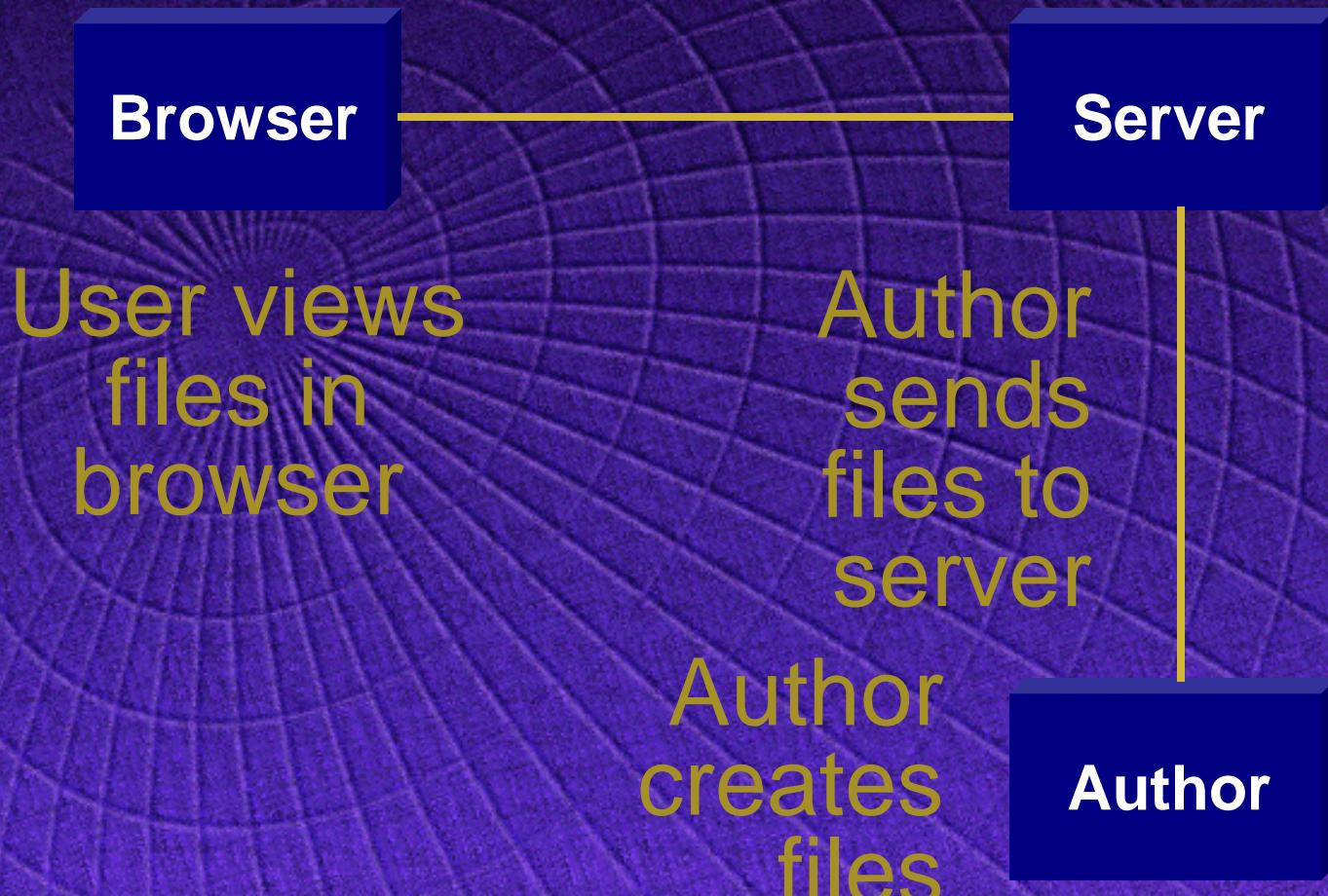
```
<H1><FONT COLOR="#b80000">  
Heading level 1</FONT></H1>  
<H2><FONT COLOR="#ff0000">  
Heading level 2 </FONT> </H2>  
<P>Paragraph with <B>bold</B> and  
<I>italic</I> text.</P>  
<HR>
```



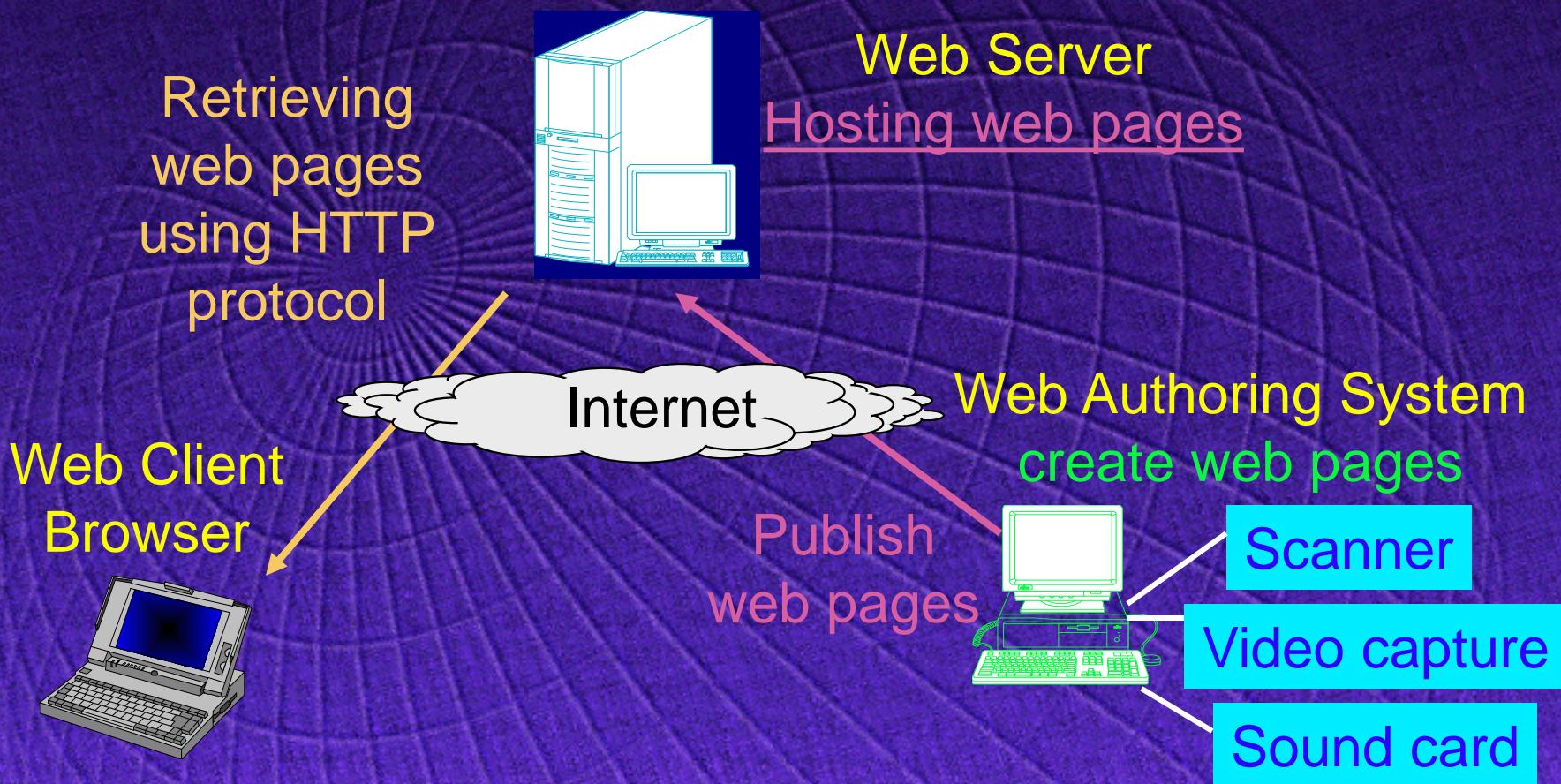
Creating HTML Files

- ◆ Text editor (Notepad, Pico)
- ◆ HTML Editor (FrontPage, Netscape Gold and HoTMetal)

Moving Files to Servers



Introduction to Client-Server Systems



Internet Client-Server Systems

Bing Maps - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Favorites Address [Go Link](http://www.bing.com/maps/default.aspx?q=%2200angji+Huang%22&mk=en-CA&FORM=BYPD#3ndoZ0MT00NzAwk0tZWxk1N0cmVdOuyYytUb3JvbnRvJTJk09udGFnLw8mc3M9e0Au/wGhbmdqa93dwFuzyU3ZD7nzdC4xJ1)

Google Web Images Videos News Maps More Symantec MEN Windows Live Sign in | Canada (English) | Extras

bing Business name or category ... Address, location, or landmark

Businesses | Collections | Locations

business results There are no business results for Xiangji Huang in current map view [Modify search](#)

welcome directions collections options share print

© 2009 Microsoft. | Privacy | Legal | Help | Feedback

Done Internet

Internet Client-Server Systems

Bing Maps - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Home

Address http://www.bing.com/maps/default.aspx?q=%22Xiangji+Huang%22&mkt=en-CA&FORM=BYFD#JndoZXJlMT00NzAwK0tIZwlk1N0cmVldCUyYytUb3JvbnnRvJTJjK09udGFyaW8mc3M9eXAuWGlhbmdqa5tIdWFuZyU3ZXNzdC4xJ1 Go Links

Google Search Bookmarks AutoFill

Web Images Videos News Maps More Sympatico / MSN Windows Live Sign in Canada (français) Extras

bing Business name or category ... Address, location, or landmark

Businesses Collections Locations

business results There are no business results for Xiangji Huang in current map view [Modify search](#)

welcome directions collections options share print

2D 3D Road Aerial Bird's eye Labels Traffic

Pan in any direction

Atkinson Rd Fine Arts Rd York Blvd

Van Macdonald Blvd York Blvd The Pond Rd

50 m

© 2009 Microsoft Corporation © 2009 NAVTEQ © AND © 2009 Pitney Bowes International Corp.

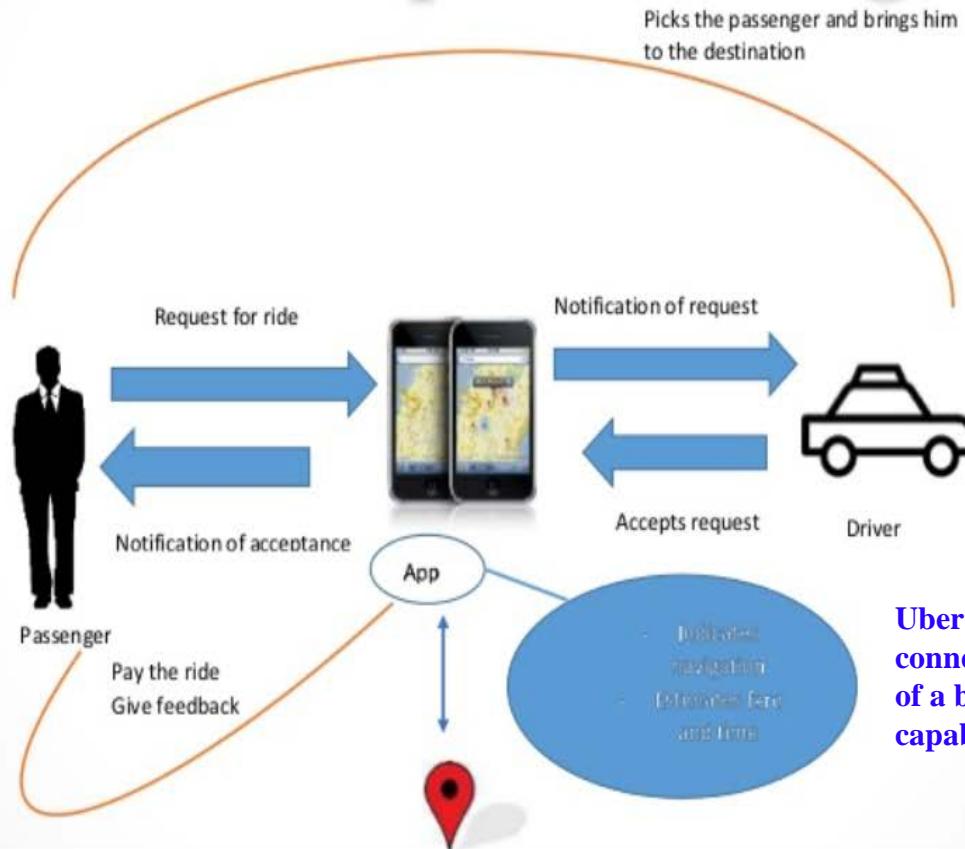
Done Internet

Internet Client-Server Systems

Presentation2 - uber

Clip slide

Order processing



Uber is an app and taxi service that connects riders and drivers with the tap of a button by using their phone's GPS capabilities

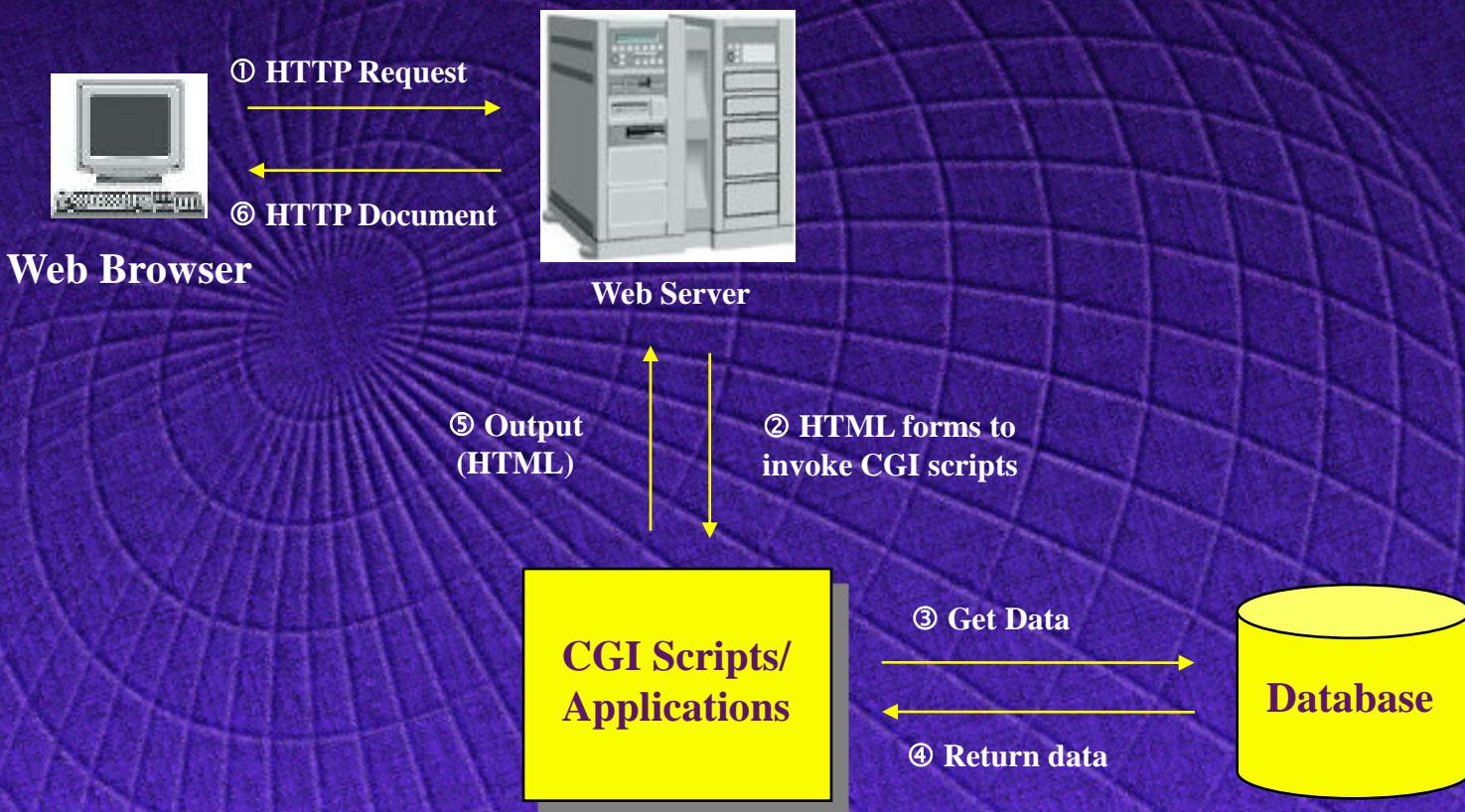
Static and Dynamic Web Pages

- ◆ A static Web page is ready before it is accessed.
- ◆ The content of a dynamic Web page is generated each time it is accessed.

Common Gateway Interface (CGI)

- ◆ CGI programming techniques were introduced to provide dynamic Web pages via server-side interaction.
- ◆ A standard method to extend the functionality of the web server.
- ◆ Any programming language can be used.
Common ones include: Perl, C++, Visual Basic.

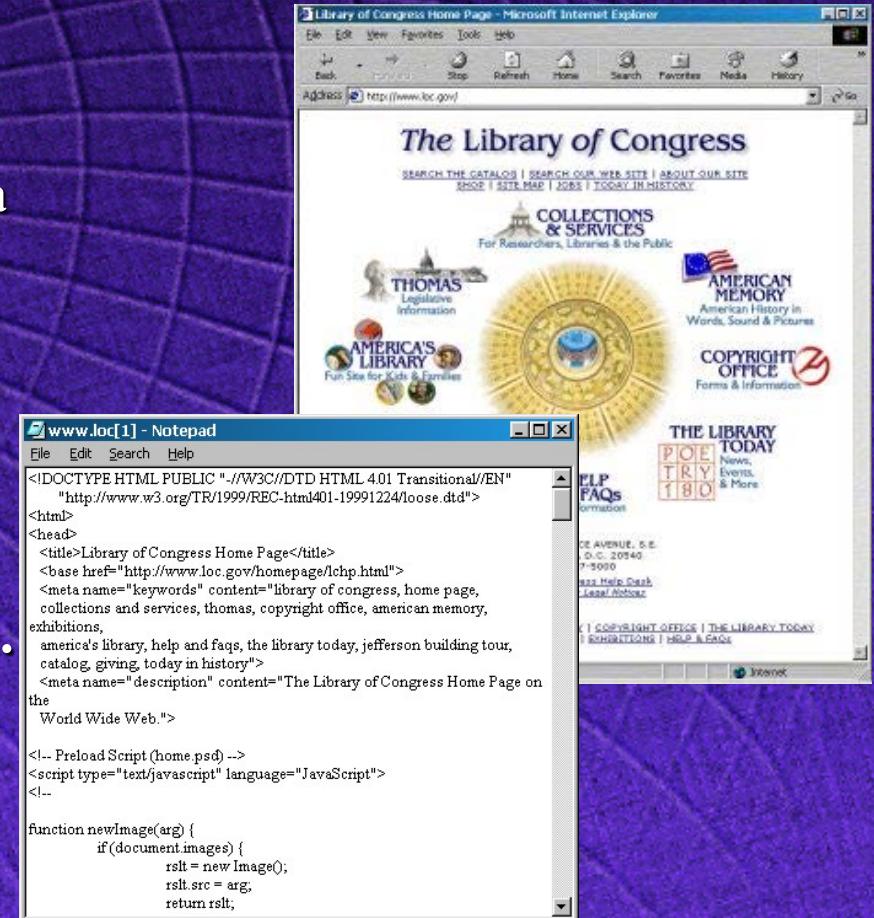
CGI-based Web Application



How Web Page Works

Sample web page and its source.

- ◆ The source contains the instructions that define the contents, layout, and structure of a web page.
- ◆ The instructions are written in HTML or another web authoring tool used in creating the page.
- ◆ The browser uses these instructions to interpret and display the web page on the screen.



How Web Page Works

URL

The screenshot shows the Microsoft Internet Explorer browser window displaying the 'ICT for Teacher Training Portal - Microsoft Internet Explorer - [Working Offline]' page. The URL 'http://www.unescobkk.org/lps/ict/ict.htm' is highlighted in red in the Address bar. The page itself features a green header with the UNESCO logo and the text 'ICT for teacher training portal'. Below the header is a navigation menu with links like 'Knowledge Resources', 'Virtual Library', 'Portals', 'Webcast', 'Electronic Articles', 'Database', 'Electronic Publications', 'Search of the Month', 'Photo Library', and 'Ordering Publications'. A search bar is also present. The main content area includes a large 'ict for teacher training' logo, a brief introduction about the portal's purpose, and three sections: '1 ICT in Education', '2 Teachers' Roles in the ICT Environment', and '3 ICT Training Strategies and Online Courses'. Each section has a thumbnail image and a brief description. At the bottom left is the UNESCO logo and the website address 'www.unescobkk.org'. The browser interface includes standard toolbar icons for Back, Forward, Stop, Refresh, Home, Search, Favorites, Media, History, Mail, Print, Edit, and Discuss.

Navigational tools

Navigational tools

Graphics /
Hyperlinks

Hyperlinks

Cookies

- ◆ A piece of information generated by the web-server and stored in the client side ready for future access.
- ◆ Cookies can make CGI scripts more interactive.
- ◆ Cookies are text files stored on Web client.
- ◆ CGI script creates cookie and has a Web server sent it to client's browser to store on hard disk.
- ◆ Later, when client revisits Web site and uses a CGI script that requests this cookie, client's browser sends information stored in the cookie.

Cookies

◆ How do cookies work?



◆ Where are cookies used?

- ✓ Shopping applications
- ✓ Storing login information
- ✓ Tracking pages visited by a user

Summary

- ◆ The Web is a networked information system that contains a huge collection of files
- ◆ The Web relies on clients and servers
- ◆ HTML and other files are sent from servers to clients
- ◆ Files are identified by URLs
- ◆ Servers send files to browsers
- ◆ Browsers interpret HTML
- ◆ Cookies is a piece of information generated by the web-server and stored in the client side.