

Chapter 16

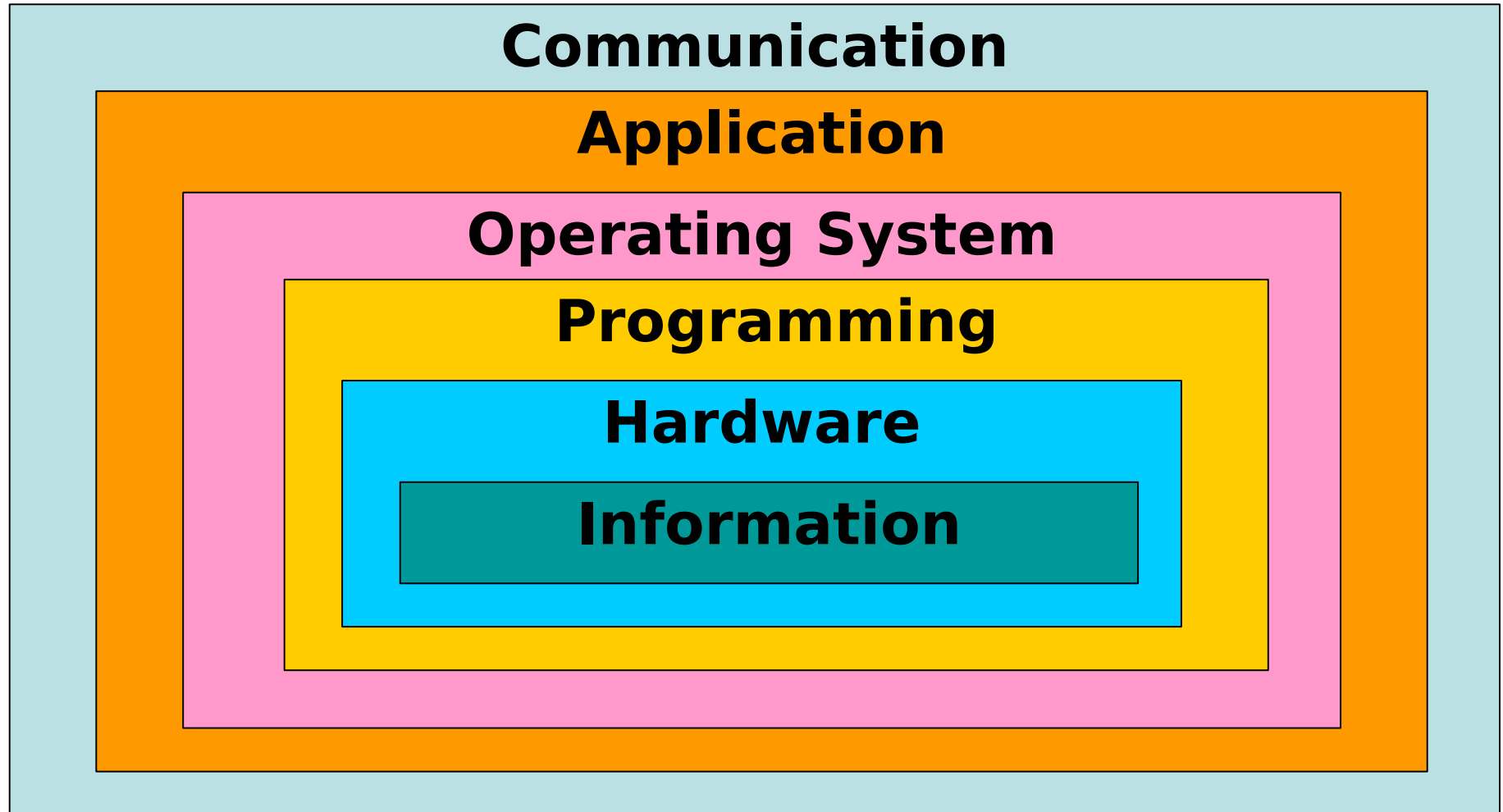
The World Wide Web



"On the Internet, nobody knows you're a dog."

The New Yorker,
Peter Steiner,
July 5, 1993

Layers of a Computing System



Chapter Goals

- Compare and contrast the **Internet** and the **World Wide Web**
- Describe general **Web processing**
- Write **basic HTML** documents
- Describe several specific **HTML tags** and their purposes

Chapter Goals

- Describe the processing of **Java applets** and **Java server pages**
- Compare and **contrast HTML** and **XML**
- Define **basic XML documents** and their corresponding DTDs
- Explain how **XML documents** are viewed

The World Wide Web

- **The Web** An infrastructure of distributed information combined with software that uses networks as a vehicle to exchange that information
- **Web page** A document that contains or references various kinds of data, such as text, images, graphics, and programs
- **Links** A connection between one web page and another that can be used “move around” as desired

Search Engines

- **Search Engine** A website that helps you find other websites
 - For example, **Snap** and **Google** are search engines
 - You enter keywords and the search engine produces a list of links to potentially useful sites
- There are two types of searches
 - Keyword searches
 - Concept-based searches

Instant Messaging

- **Instant messaging** (IM) An application that allows people to send and receive messages in real time
 - Both sender and receiver must have an IM running
 - Most IM applications use a proprietary protocol that dictates the precise format and structure of the messages that are sent across the network to the receiver.
 - Instant messages are not secure

Cookies

- **Cookie** A small text file that a web server stores on your local computer's hard disk
 - A cookie contains information about your visit to the site
 - Cookies can be used
 - to determine number of unique visitors to the site
 - to customize the site for your future visits
 - to implement shopping carts that can be maintained from visit to visit
 - Cookies are not dangerous

Web Browser

- **Browser** A software tool that issues the request for the web page we want and displays it when it arrives
- We often talk about “visiting” a website, as if we were going there
 - In truth, we actually specify the information we want, and it is brought to us
 - The concept of visiting a site is understandable in that we often don’t know what’s at a particular site until we “go to it” and see

Web Browser

- **Web server** The computer that is set up to respond to web requests
- **Web address** The core part of a **Uniform Resource Locator, or URL,** which uniquely identifies the page you want out of all of the pages stored anywhere in the world

Web Browser

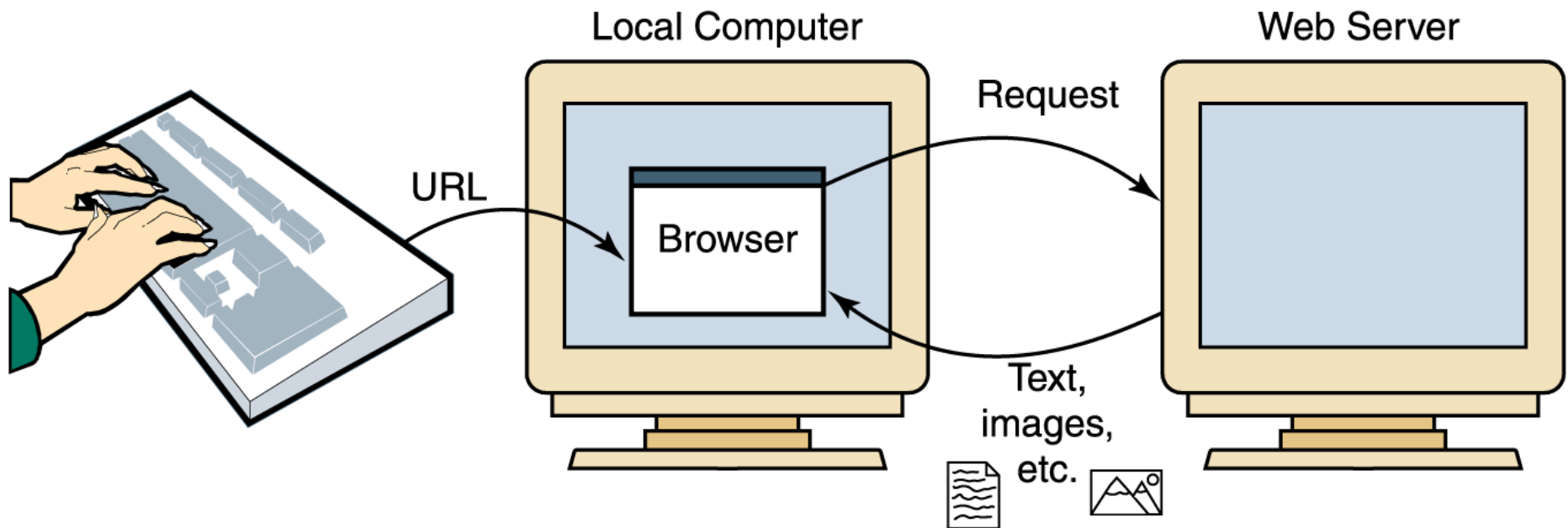


Figure 16.2 A browser retrieving a Web page

Uniform Resource Identifiers

- Uniform naming schema to identify **resources** on the Internet
- A resource can be anything:
 - `index.html`
 - `mysong.mp3`
 - `picture.jpg`

- Example URIs:

`http://www.cs.wisc.edu/~dbbook/index.html`

`mailto:webmaster@bookstore.com`

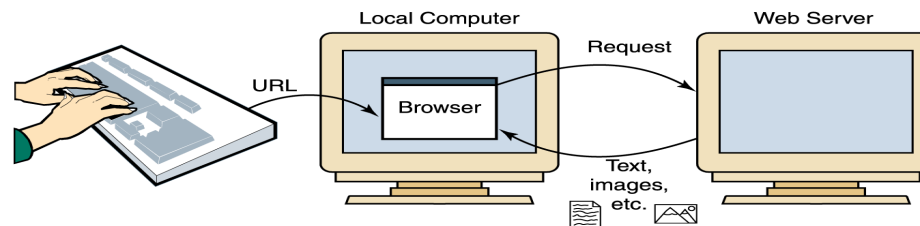
Structure of URIs

`http://www.cs.wisc.edu/~dbbook/index.html`

- URI has three parts:
 - **Naming schema** (http)
 - **Name of the host computer** (`www.cs.wisc.edu`)
 - **Name of the resource** (~dbbook/index.html)
- URLs are a subset of URIs

Hypertext Transfer Protocol

- ◆ What is a communication protocol?
 - Set of standards that defines the structure of messages
 - Examples: TCP, IP, **HTTP**
- ◆ What happens if you click on `www.cs.wisc.edu/~dbbook/index.html?`
 - Client (web browser) sends HTTP request to server
 - Server receives request and replies
 - Client receives reply; makes new requests



HTTP (Contd.)

Client to Server:

```
GET ~/index.html HTTP/1.1
User-agent: Mozilla/4.0
Accept: text/html, image/gif,
        image/jpeg
```

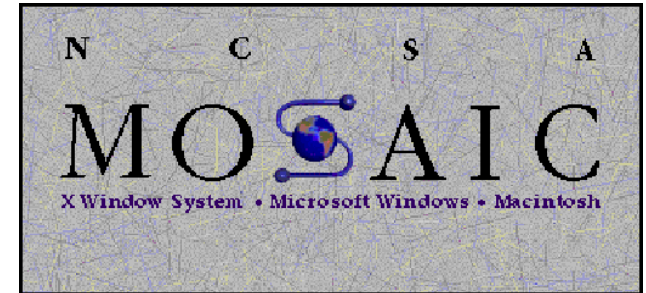
Server replies:

```
HTTP/1.1 200 OK
Date: Mon, 04 Mar 2002 12:00:00
      GMT
Server: Apache/1.3.0 (Linux)
Last-Modified: Mon, 01 Mar 2002
      09:23:24 GMT
Content-Length: 1024
Content-Type: text/html
<HTML> <HEAD></HEAD>
<BODY>
<h1>Barns and Nobble Internet
      Bookstore</h1>
Our inventory:
<h3>Science</h3>
<b>The Character of Physical
      Law</b>
```

...

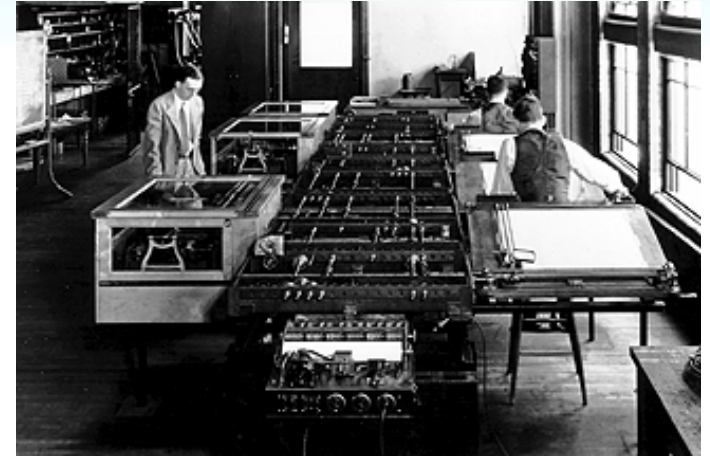
A Simple Idea

- **Marc Andreessen**, born in Iowa, 1971, received his BA in computer science from the University of Illinois, Urbana-Champaign, where research was being done at the university's **National Center for Supercomputing Applications** on early browsers such as **ViolaWWW** (created by **Pei-Yuan Wei** in 1993), based on **Tim Berners-Lee's** open standards for the **World Wide Web**.
- These earlier browsers had been created to work only on expensive Unix workstations, so Andreessen and a full-time salaried co-worker Eric Bina worked on creating an improved and user-friendlier version with integrated graphics that would work on personal computers. The resulting code was the Mosaic web browser.
- After graduating, he teamed up with with **Jim Clark**, the founder of **Silicon Graphics**, and formed the **Mosaic Communications Corporation** in Mountain View, California, with Andreessen appointed as a vice-president. The University of Illinois was unhappy with the use of the Mosaic name, so Mosaic Communications changed its name to **Netscape Communications**, and its flagship web browser was the **Netscape Navigator**.
- <http://www.ibiblio.org/pioneers/andreesen.html>



A Previous Simple Idea

- **Vannevar Bush**, born 1890, in Chelsea, Massachusetts, was never directly involved with the creation or development of the Internet. Yet many consider Bush to be the Godfather of our wired age often making reference to his 1945 essay, "**As We May Think.**"
- In his article, Bush described a theoretical machine he called a "**memex,**" which was to enhance human memory by allowing the user to store and retrieve documents linked by **associations** (what he called **associative trails**). This associative linking was very similar to what is known today as **hypertext.**
- Vannevar Bush died on June 30, 1974, years before the Internet became widely popular or the World Wide Web even existed. With the growing popularity of the Internet many now look back through its history and see Bush as a visionary.
- **<http://www.ibiblio.org/pioneers/bush.html>**



Differential Analyzer - 1931



Tim Berners-Lee



Biography

<http://www.ibiblio.org/pioneers/lee.html>

<http://www.w3.org/People/Berners-Lee/>

IBM DeveloperWorks Interview

<http://www-128.ibm.com/developerworks/podcast/dwi/cm-int082206.html>

HTML

- Web pages are created (or built) using a language called the **Hypertext Markup Language**, or **HTML**
- The term **markup language** comes from the fact that the primary elements of the language take the form of **tags** that we insert into a document to annotate the information stored there

HTML

Final Report

European Conference on Expert Systems

boldface

Submitted by Justin Parker

Center

First of all, our thanks go out to the following sponsors for their support of the conference and its supplemental activities.

Allied Interactive

Sybernetics, Inc.

Dynamic Solutions of New Jersey

make these bullets

The conference was a great success. It ran a full four days, including workshops and special sessions. Subjective feedback from conference attendees was largely positive, and financially the revenues resulted in a surplus of over \$10,000.

Figure 16.2

A marked-up document

HTML

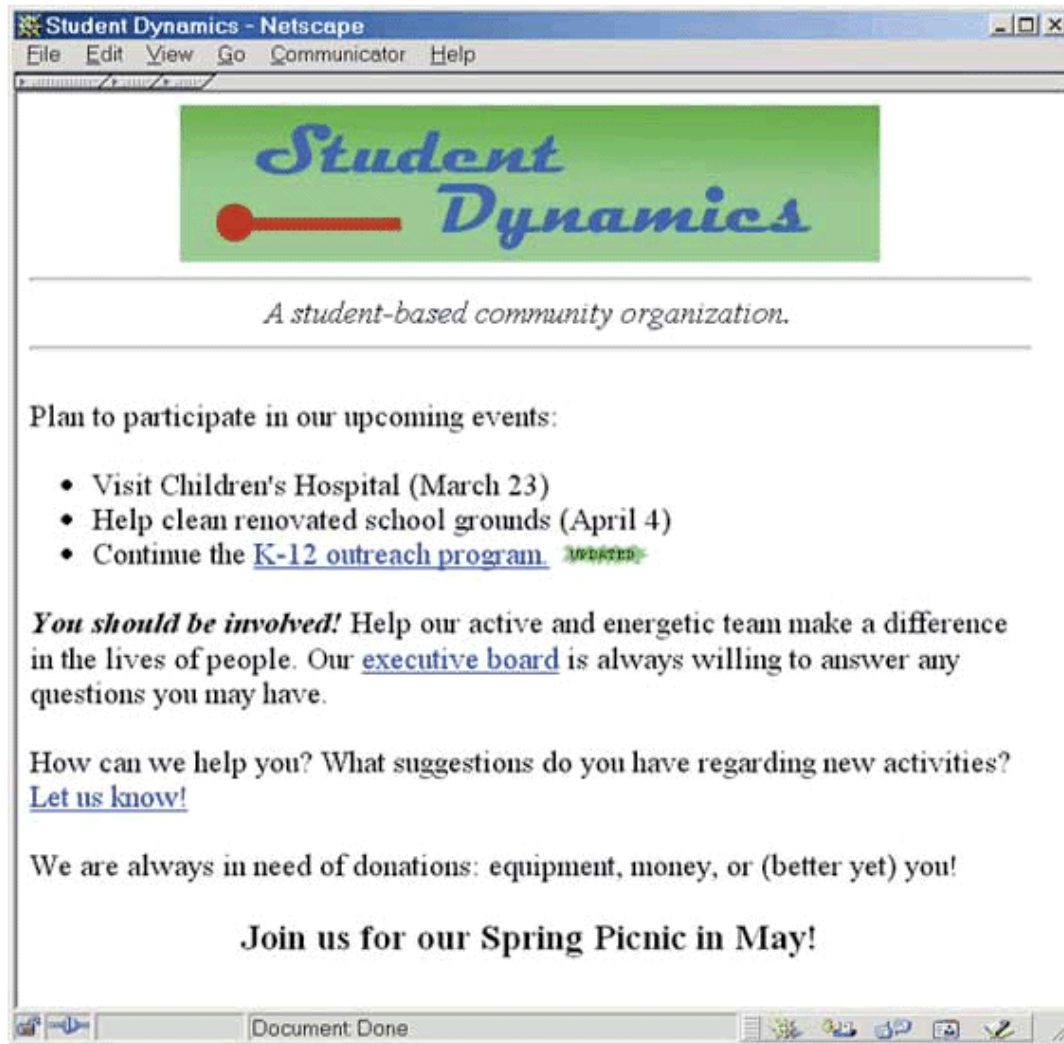


Figure 16.3
The Student Dynamics Web page as displayed in Netscape Navigator

HTML

```
<HTML>
  <HEAD>
    <TITLE>Student Dynamics</TITLE>
  </HEAD>
  <BODY>
    <CENTER><IMG SRC="stuDynamics.gif"></CENTER>
    <HR>
    <CENTER><I>A student-based community organization.</I></CENTER>
    <HR>
    <P>Plan to participate in our upcoming events:</P>
    <UL>
      <LI>Visit Children's Hospital (March 23)</LI>
      <LI>Help clean renovated school grounds (April 4)</LI>
      <LI>Continue the <a href="outreach.html">K-12 outreach
        program.</a> <IMG SRC="updated.gif"></LI>
    </UL>
    <P><B><I>You should be involved!</B></I> Help our active and
      energetic team make a difference in the lives of people. Our
      <a href="execBoard.html">executive board</a> is always willing
      to answer any questions you may have.</P>
    <P>How can we help you? What suggestions do you have regarding
      new activities? <a href="suggestions.html">Let us know!</a></P>
    <P>We are always in need of donations: equipment, money, or
      (better yet) you!</P>
    <CENTER><H3>Join us for our Spring Picnic in May!</H3></CENTER>
  </BODY>
</HTML>
```

Figure 16.4
The HTML document
defining the Student
Dynamics Web page

HTML

- **Tags** are enclosed in angle brackets (`< . . . >`)
- Words such as HEAD, TITLE, and BODY are called **elements** and specify the type of the tag
- Tags are often **used in pairs**, with a start tag such as `<BODY>` and a corresponding end tag with a / before the element name, such as `</BODY>`

HTML

- The **browser** determines **how** the page should be **displayed** based on the tags
- The browser
 - Ignores the way we format the HTML document using carriage returns, extra spaces, and blank lines
 - Takes into account the width and height of the browser window
 - Reformats the contents to fit your browser window

Basic HTML Formatting

- The **paragraph tags** (`<P> . . . </P>`) specify text that should be treated as a separate paragraph
- The **center tags** (`<CENTER> . . . </CENTER>`) indicate that the enclosed information should be centered in the browser window

HTML Fundamentals

- `<h1>`An important heading`</h1>`
- `<h2>`A slightly less important heading`</h2>`
- `<p>`This is the first paragraph.`</p>`
`<p>`This is the second paragraph.`</p>`
- This is a really ``**interesting**`` topic!

HTML Fundamentals

An important heading

A slightly less important heading

This is the first paragraph.

This is the second paragraph.

This is a really *interesting* topic!

Basic HTML Formatting

- The B, I, and U elements are used to indicate that the enclosed text should be **bold**, *italic*, or underlined, respectively
- The <HR> tag inserts a **horizontal rule** (that is, a line) across the page

Basic HTML Formatting

- We often have cause to display a **list** of items
 - The UL element stands for an **unordered list**, and the LI element represents a **list item**
- Several elements are used to define headings in a document
 - There are six predefined **heading** elements defined in HTML: H1, H2, H3, H4, H5, and H6

Images and Links

- Many tags can contain attributes that indicate additional details about the information or how the enclosed information should be displayed
 - An **image** can be incorporated into a web page using the IMG element, which takes an attribute that identifies the image file to display
 - ``

Images and Links (cont.)

- A **link** is specified using the element A, which stands for anchor
- The tag includes an attribute called **HREF** that specifies the URL of the destination document.

For example

```
<A HREF = "http://duke.csc.villanova.edu/docs/">  
Documentation Central!</A>
```

The Basics of HTML

- <http://www.w3.org/MarkUp/Guide/> - Getting started with HTML
- <http://www.w3.org/MarkUp/Guide/Advanced.html> - More advanced features
- <http://www.w3.org/MarkUp/Guide/Style.html> - Adding a touch of style
- <http://www.w3.org/People/Raggett/tidy/> - HTMLTidy
- <http://validator.w3.org/> - Ensures proper HTML code

Other Useful Websites

- <http://www.sitepoint.com/> - lots of helpful information and excellent books
- <http://www.csszengarden.com/> - css Zen Garden
- <http://www.w3.org/XML/> - everything you ever wanted to know about XML
- <http://www.w3.org/Style/XSL/> - definitive source for XSL

Interactive Web Pages

- When HTML was first developed, there was **no way to interact** with the information and pictures presented in a web page
- As users have clamoured for a **more dynamic web**, new technologies were developed to accommodate these requests
- Many of the new ideas were offshoots of the newly developed **Java** programming language

Java Applets

- **Java applet** A program that is designed to be embedded into an HTML document and transferred over the Web to someone who wants to run the program

An applet is embedded into an HTML document using the APPLET tag

```
<APPLET code="MyApplet.class" width=250  
height=150 ></APPLET>
```

Java Applets

- A browser has a **built-in interpreter** that executes the applet, allowing the user to interact with it.
 - Consider the difficulties inherent in this situation
 - How can we execute a program that was written on one type of computer on possibly many other types of computers?

Java Applets

- Java programs are compiled into **Bytecode**, a low-level representation of a program that is not the machine code for any particular type of CPU
- Java applets are restricted as to what they can do
 - The Java language has a carefully constructed **security** model
 - An applet, for instance, cannot access any local files or change any system settings

Java Server Pages

- A Java Server Page, or JSP, is a web page that has **JSP scriptlets** embedded in them
- **Scriptlet** A small piece of executable code intertwined among regular HTML content
- Also called a **Servlet!**

Java Server Pages

- A JSP scriptlet is encased in special tags beginning with `<%` and ending with `%>`
- Imagine JSP scriptlets as having the expressive power of a full programming language

```
<H3>  
<%  
out.println ("hello there");  
%>  
</H3>
```

Java Server Pages

- Note that JSPs are **executed on the server side** where the web page resides
- By the time it arrives at your computer, all active processing has taken place, producing a static (though **dynamically created**) web page
- JSPs are particularly good for coordinating the **interaction** between a web page and an **underlying database**

XML

- **HTML** has a **predefined set of tags** and each tag has its own meaning
- There is nothing about HTML tags that describes the true content of a document
- The **Extensible Markup Language**, or **XML**, allows the creator of a document to describe its contents by **defining** his or her **own set of tags**

XML

- **Metalinguage** A language for talking about, or defining, other languages
- XML is a metalanguage

XML

```
<?xml version="1.0" ?>
<!DOCTYPE books SYSTEM "books.dtd">
<books>
  <book>
    <title>The Hobbit</title>
    <authors>
      <author>J.R.R. Tolkien</author>
    </authors>
    <publisher>Ballantine</publisher>
    <pages>287</pages>
    <isbn>0-345-27257-9</isbn>
    <price currency="USD">7.95</price>
  </book>
  <book>
    <title>A Beginner's Guide to Bass Fishing</title>
    <authors>
      <author>J. T. Angler</author>
      <author>Ross G. Clearwater</author>
    </authors>
    <publisher>Quantas Publishing</publisher>
    <pages>750</pages>
    <isbn>0-781-40211-7</isbn>
    <price currency="USD">24.00</price>
  </book>
</books>
```

- Like HTML, an XML document is made up of tagged data

Figure 16.5 An XML document containing data about books

XML

- **Document Type Definition (DTD)** A specification of the organization of the document
- The structure of a particular XML document is described by its corresponding DTD document

```
<!ELEMENT books (book*) >
<!ELEMENT book (title, authors, publisher, pages, isbn, price)>
<!ELEMENT authors (author+)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT author (#PCDATA)>
<!ELEMENT publisher (#PCDATA)>
<!ELEMENT pages (#PCDATA)>
<!ELEMENT isbn (#PCDATA)>
<!ELEMENT price (#PCDATA)>
<!ATTLIST price currency CDATA #REQUIRED>
```

Figure 16.6 The DTD document corresponding to the XML books document

XML

- XML represents a standard format for organizing data without tying it to any particular type of output
- **Extensible Stylesheet Language (or XSL)** A language for defining transformations from XML documents to other output formats

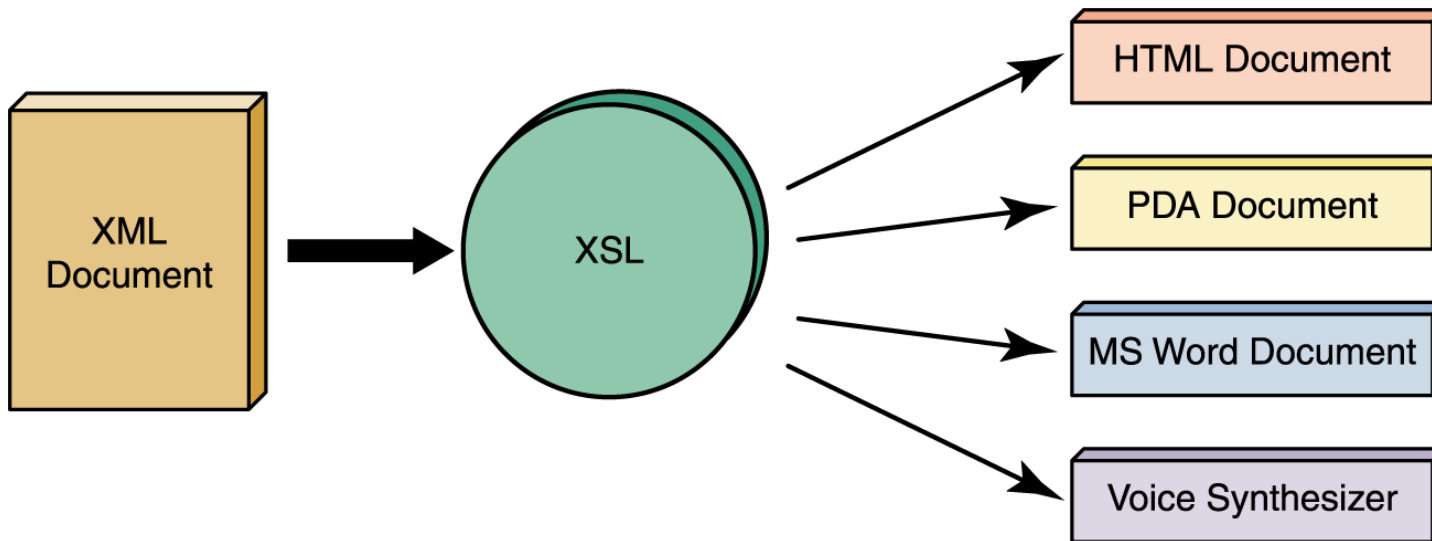
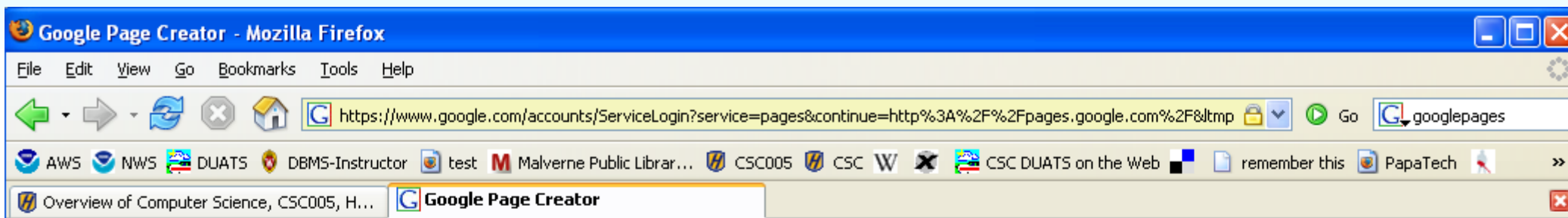


Figure 16.7
An XML document can be transformed into many output formats

http://pages.google.com



Welcome to Google Page Creator

Create your own web pages, quickly and easily.

Google Page Creator is a free online tool that makes it easy for anyone to create and publish useful, attractive web pages in just minutes.

- **No technical knowledge required.**

Build high-quality web pages without having to learn HTML or use complex software.

- **What you see is what you'll get.**

Edit your pages right in your browser, seeing exactly how your finished product will look every step along the way.

- **Don't worry about hosting.**

Your web pages will live on your own site at <http://yourgmailusername.googlepages.com>

Google Page Creator is a [Google Labs](#) project, and is still in an early testing phase. If you're interested in taking it for a test drive, login with your Gmail account to begin making pages. If you don't have a Gmail account and you have a mobile phone, you can [sign up here](#). We invite you to let us know what you think by sending us your [feedback and suggestions](#).

Sign in to Google Page Creator with your

Google Account

Email:

Password:

Remember me on this computer.

[Forgot your password?](#)

[No Gmail Account? Sign up](#)

[About Google Page Creator](#)

Previous Homework

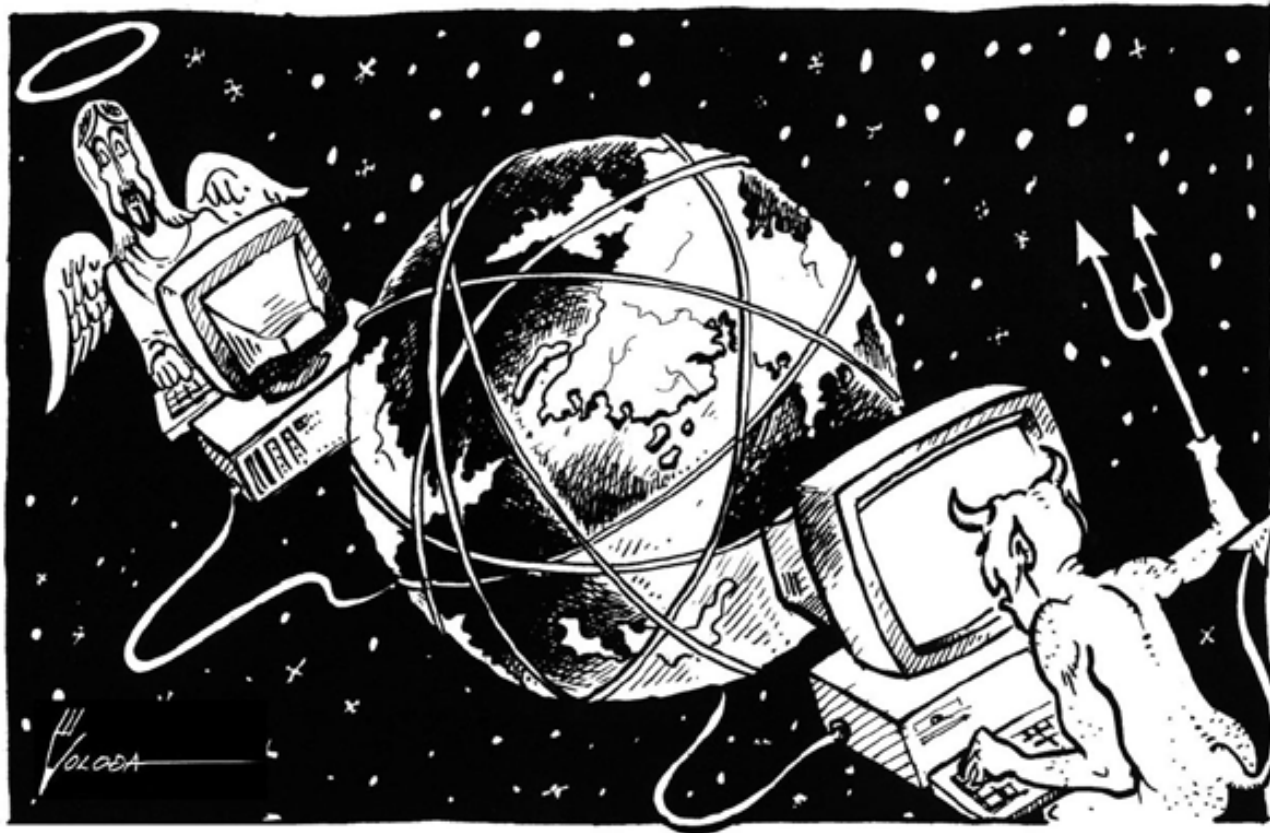
- Get a gmail account...
- ...if you don't want to use your mobile (or you don't have one), send an email to:
papacosta@gmail.com
- When you get an account, send me a message!
papacosta@gmail.com
- **This becomes our 1st Assignment!**

Assignment One

- Create A **Simple** Home Page – 6/10 pts
- Create A **Complex** Home Page – 8/10 pts
 - Links, images, color fonts, multi-column layouts
- Create A **Multilevel Site** – 10/10 pts
 - Tree structure
 - Multiple pages
- **Due Next Wednesday**
 - No lateness

Homework

- **Read Chapter Sixteen**



...be careful on the web!