

SESSION: Web and XML

- Basic Web

- HTML
- Web browsers
- Web servers and proxies
- HTTP

- XML and XML Family

- Semantic Web: The Next Generation Web

- Object-orienting the Web

- Quick overview of XML-Web services

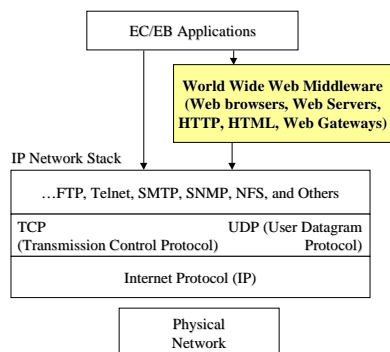
Anjad Umar

World Wide Web

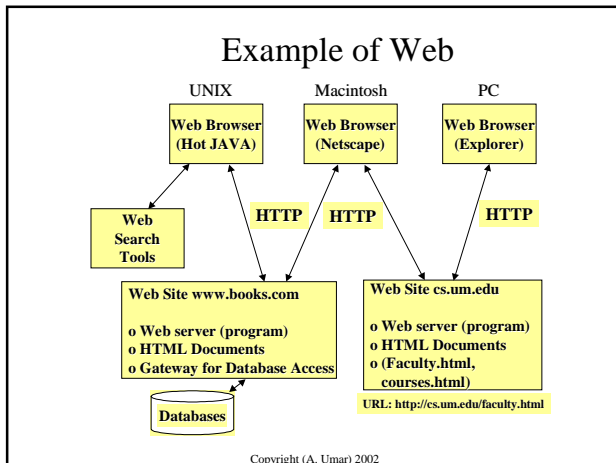
- WWW is a collection of middleware that operates on top of the Internet.
- WWW middleware supports the growing number of users and applications
- Basic WWW middleware is based on a few simple concepts and technologies
 - Web servers
 - Web browsers
 - Uniform Resource Locator (URL)
 - Hypertext Transfer protocol (HTTP)
 - Hypertext Markup Language (HTML)
 - Gateways to non-Web resources

Copyright (A. Umar) 2002

World Wide Web Overview



Copyright (A. Umar) 2002



Web Browsers

- Web browsers are clients designed to display information prepared in a markup language, known as HTML.
- Common Web browsers (Netscape, Microsoft)
- Web browser uses HTTP to fetch the needed document from an appropriate Web server.
- HTTP compliance is a basic requirement for Web browsers
- Basic browsers are relatively dumb (they just pass user requests to Web servers and display the results)
- Java applets can run on Java compatible browsers
- Java applets are downloaded to the Java enabled browsers where they run producing graphs/charts, invoking multimedia applications, etc

Copyright (A. Umar) 2002

Hypertext Markup Language (HTML)

- Easy to use language that tags the text files for display at Web browsers.
- Also helps in creation of hypertext links (called hyperlinks)
- Hyperlinks contain URLs for the needed resources.
- HTML is not WYSIWYG
- HTML contains instructions for highlights, bullets, etc.
- HTML documents can be created in text or through HTML editors (can be saved as from Word, etc)
- Transformers available to convert word processing documents to HTML (e.g., Word, Framemaker)
- How to convert database information to HTML

Copyright (A. Umar) 2002

Consulting Group1

Welcome to our consulting group. You can do the following:

- Read about our services
- Access home pages of the groups we work with (WWW)

Now choose the connections by clicking on the following

- [Our services](#)
- [Web Information](#)

The following HTML statements can be used to design this home page (we have inserted appropriate URL's for the hot links):

```
<html>
<TITLE>Consulting Group1 </TITLE>
<H1>Consulting Group1 </H1>
<P> Welcome to our consulting group. You can do the following:
<UL>
<LI> Read about our services .
<LI> Access home pages of the groups we work with (WWW)
</UL>
<P> Now choose the connections by clicking on the following
<a href="http://www.myserver.com/services.html"> Our services</a>
<a href="http://www.w3.org"> Web Information</a>
</html>
```

Copyright (A. Umar) 2002

Common HTML Tags

<H1> Heading </H1>

<P>

 WWW Information

<APPLET CODE=myapplet.class WIDTH=110 HEIGHT=100>

</APPLET>

– Imbeds a Java applet (myapplet) in HTML page

– The applet code has to reside on the same server as HTML pages

<FORM ACTION="/cgi.bin/viewguery">

– Invokes a cgi program (viewguery) on the server

– cgi programs are typically written in scripting languages (e.g., PERL) but can be written in any language (C, C++, Java)

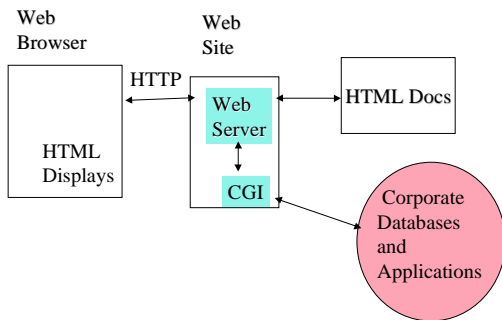
Copyright (A. Umar) 2002

Web Servers

- Web servers provide the content for Web users.
- Web servers are populated by the content providers.
- Conceptually: a Web server is a catalog of information
- In reality: a Web server is
 - Server software (e.g., Apache)
 - a collection of HTML files
 - Gateways to non-HTML resources (CGI, Servlet,,)
- In many cases, a machine is dedicated/designated as a Web server .
- Convention: content providers begin with designated "home pages"
- Home pages include company logo, fancy artwork for attention, special deals, overviews, pointers to additional information, etc.

Copyright (A. Umar) 2002

First Generation Web Architectures (HTML, HTTP, CGI)



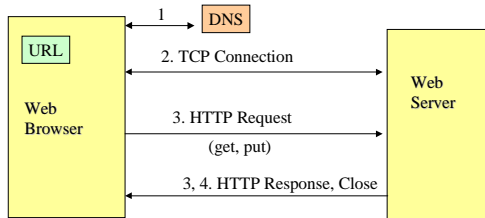
Copyright (A. Umar) 2002

Hypertext Transfer Protocol (HTTP)

- Hypertext Transfer Protocol (HTTP) is an application-level protocol designed for Web users.
- Global URI, e.g., <http://www.upenn.edu> (uses DNS, no need for a new directory)
- Request/response Model. HTTP uses an extremely simple request/response model :
 - Establishes connection with the Web server specified in the URL
 - Retrieves (get) or updates (put, post, delete) the needed document (typically form input, e.g., card number)
 - Closes the connection.
- Statelessness: Every time you click on a hyperlink, you are initiating an HTTP session

Copyright (A. Umar) 2002

Web Interactions



Basic Flow (LOTIC):

1. Locate (find) the partner
2. Open (connect, bind)
3. Transmit (Request/Response)
4. Close

HTTP 1.0 Methods:
Initial (official)

- Get
- Head
- Post

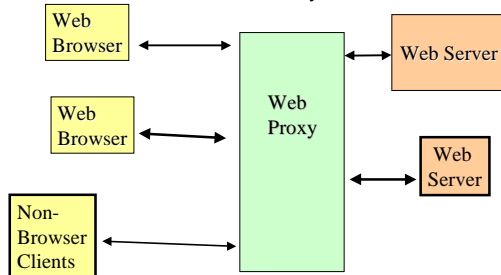
Added:

- Put
- Delete
- Link/Unlink

Copyright (A. Umar) 2002

Web Proxies

- Proxies sit between clients and servers
- Act as a server to client and vice versa
- can be used for security, workload balance, etc



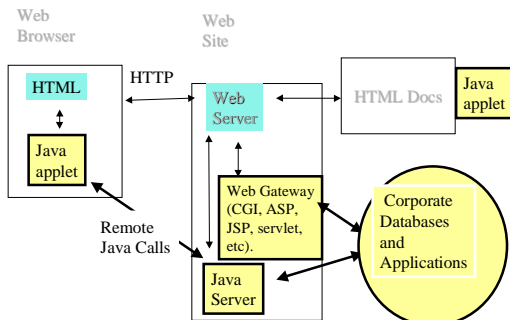
Copyright (A. Umar) 2002

Gateways to non-Web resources

- The gateway programs typically run in Web servers.
 - Common Gateway Interface (CGI) is the oldest technology
 - Server pages: ASP, JSP, XSP (*SP)
 - Servlets are becoming popular
- Java applets can provide gateway functionality
- "Relational gateways" for relational databases from Web browsers are COTS
- Other Web gateways are evolving for Web-based architectures

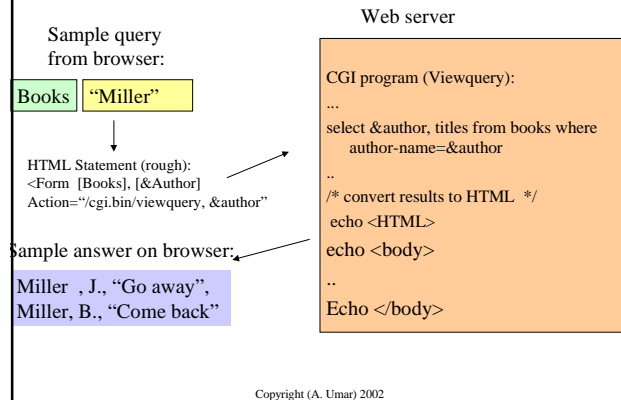
Copyright (A. Umar) 2002

Evolving Web Architectures

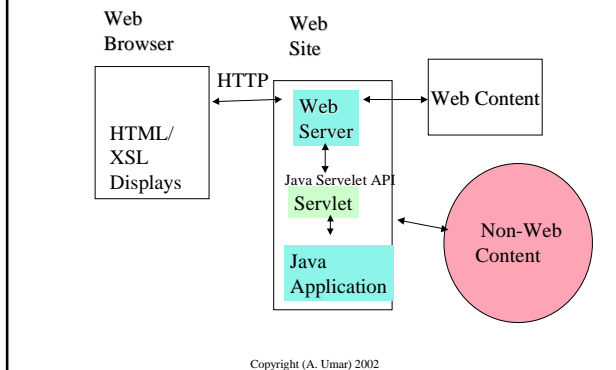


Copyright (A. Umar) 2002

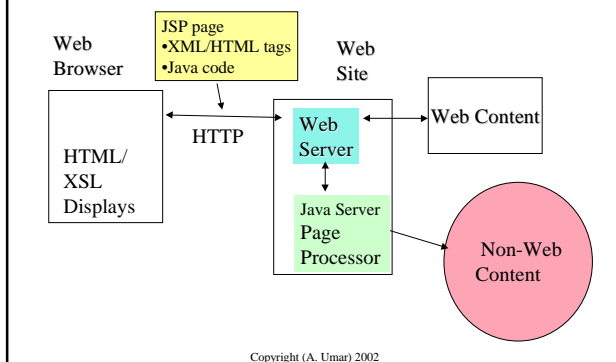
Example of CGI (Simplified View)



Servlets



JSPs



XML (eXtensible Markup Language)

- Gaining importance for common data representation
- “Simpler” than SGML (subset of SGML)
- More “general” than HTML
- Example:

```
<CUSTOMER>
  <NAME> Joe </NAME>
  <ADDRESS> NY </ADDRESS>
</CUSTOMER>
```
- Great deal of activity in Ecommerce (competition to EDI), messaging middleware, data transformers, data management, publishing, etc.

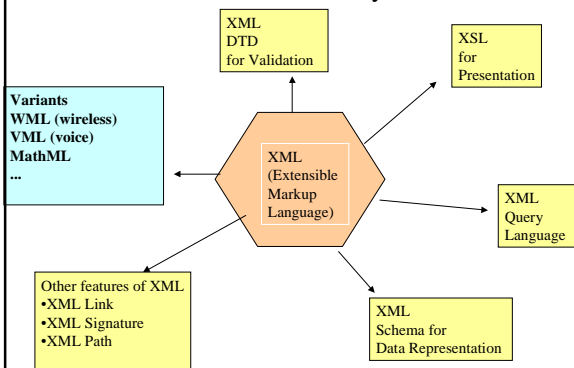
Copyright (A. Umar) 2002

XML: How is it used

- Web
 - New browsers starting to support XML
- Electronic commerce
 - Possible replacement for EDI (Electronic Data Interchange)
- Data management
 - XMI: XML Metadata Interchange, used to exchange UML models between vendor tools
 - CWM: Common Warehouse Metadata (Oracle + others)
- Publishing
 - Often used in place of SGML because of its “lightness”
- Application to application message format

Copyright (A. Umar) 2002

XML Family



Copyright (A. Umar) 2002

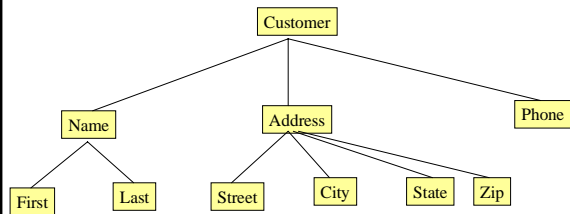
Customer Record in XML

```
<?xml version="1.0" standalone="yes"?>
<customer>
  <name>
    <first>Pat</first>
    <last>Hemsath</last>
  </name>
  <address>
    <street>Dead-end Building </street>
    <street>Horror Street</street>

    <city>Morristown</city><state>NJ</state>
    <zip>07960</zip>
  </address>
  <phone>973-555-3114</phone>
</customer>
```

Copyright (A. Umar) 2002

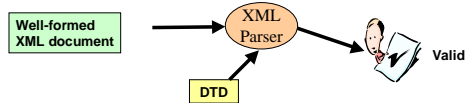
XML Tree



Copyright (A. Umar) 2002

Well-formed versus Valid

- A Well-Formed Document:
 - adheres to the syntactic rules defined by the XML standard
 - E.g. Tags are delimited by < and >
- A Valid Document:
 - A well-formed document that also adheres to the rules of a specified Document Type Definition (DTD)
- DTD:
 - specifies a set of rules for the structure of the document



Copyright (A. Umar) 2002

A DTD for a Customer

```
<!ELEMENT customer (name, address?, phone?)>
<!ATTLIST customer id CDATA #REQUIRED>
<!ELEMENT name (first, middle?, last)>
<!ELEMENT address (street+, city, state, zip)>
<!ELEMENT phone (#PCDATA)>
<!ELEMENT first (#PCDATA)>
<!ELEMENT middle (#PCDATA)>
<!ELEMENT last (#PCDATA)>
<!ELEMENT street (#PCDATA)>
<!ELEMENT city (#PCDATA)>
<!ELEMENT state (#PCDATA)>
<!ELEMENT zip (#PCDATA)>
```

... We can save this into a file called *customer.dtd*

Copyright (A. Umar) 2002

Valid XML document

```
<?xml version="1.0" standalone="no"?>
<!DOCTYPE customer SYSTEM "customer.dtd">
<customer id="12345">
  <name>
    <first>Joe</first>
    <last>Bubba</last>
  </name>
  <address>
    <street>Building2</street>
    <street>445 Swamp Street</street>
    <city>Joytown</city>
    <state>NJ</state>
    <zip>07555</zip>
  </address>
  <phone>933-555-1111</phone>
</customer>
```

Copyright (A. Umar) 2002

Well-formed but *not* Valid XML document

```
<?xml version="1.0" standalone="no"?>
<!DOCTYPE customer SYSTEM "customer.dtd">
<fool id="12345">
  <foo2>
    <foo3>Sam</foo3>
    <foo4>Sheppard</foo4>
  </foo2>
  <foo5>
    <foo6>Office</foo6>
    <foo7>445 Street</foo7>
    <foo8>Hilltown</foo8>
    <foo9>MI</foo9>
    <foo10>48190</foo10>
  </foo5>
  <foo11>734-555-2222</foo1>
</fool>
```

Copyright (A. Umar) 2002

XML: What is it composed of

- Elements and sub-elements
 - Attributes
- Entity References
 - Used to represent special characters
- Comments
 - Comments begin with `<!--` and end with `-->`.
- Processing Instructions
 - Escape hatch to provide information to an application
- CDATA sections
 - Instructs the parser to ignore most markup characters
- Document Type Declarations (DTD)
 - Required in SGML, optional in XML

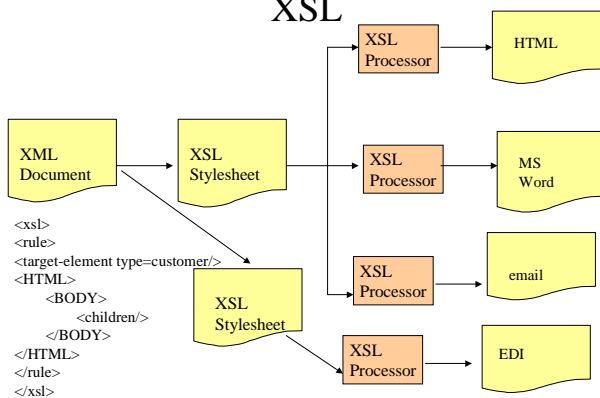
Copyright (A. Umar) 2002

XML: What is it composed of

- DTD's
 - Allows a document to communicate meta-information to parser about its content
- DTD's contain:
 - Element Type Declarations
 - Identify names of elements and nature of their content.
 - `<!ELEMENT customer (name, address)>`
 - `<!ELEMENT address (type, street+, city, state, zip)>`
 - Attribute List Declarations
 - Identify which elements may have attributes, what attributes they may have, what values the attributes may hold, and what value is the default
 - Entity Declarations
 - Allow you to associate a name with some other fragment of content

Copyright (A. Umar) 2002

XSL



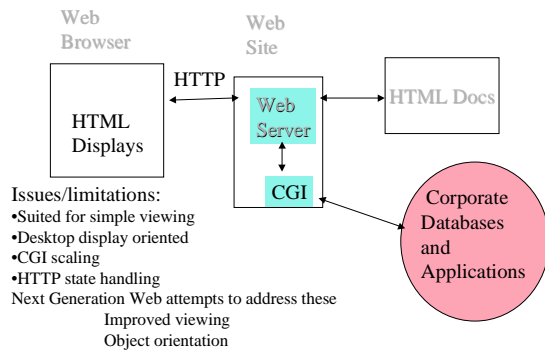
Copyright (A. Umar) 2002

Semantic Web - Next Generation Web

- Key idea: expanding the scope of web from human interactions to machine interactions
- HTML to XML : content parsing
- HTTP to HTTP-NG: State handling
- Improved programmability:
 - Access remote objects
 - Object view (Document Object Model)
 - Many others (detailed programmability, digital signatures)
- Reference: www.w3.org

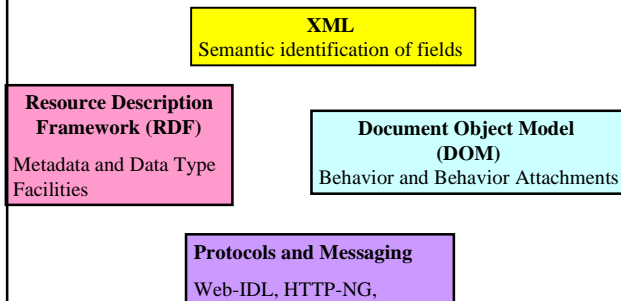
Copyright (A. Umar) 2002

First Generation Web Architectures

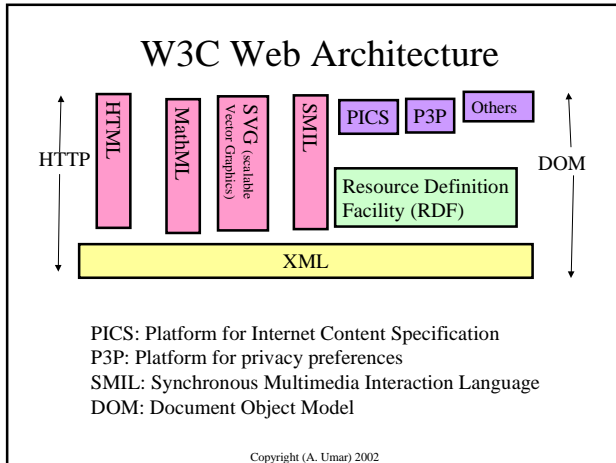


Copyright (A. Umar) 2002

Next Generation Web Model - Key Players



Copyright (A. Umar) 2002



Extensions for Viewing

- XHTML - Combines XML and HTML
- Voice browsers (workshop 1998) - voice markup language
- Web mobile access (Workshop 1998) - working closely with WAP
- Synchronized multimedia integration language (SMIL) for presentation

Copyright (A. Umar) 2002

Resource Description Framework

- Uses XML to define metadata for Web
- Example:

```
<RDF xmlns="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:DC="http://purl.org/dc/elements/1.0/"
  <Description about = "http://www.telcordia.com/announce1.html"
  <DC:title> Telcordia Personnel </DC:title>
  <DC:date> 1999-10-10 </DC:date>
  <DC:subject> New benefits </DC:subject>
  </description>
</RDF>
```
- Name conflicts resolved through XML Namespaces
- RDF can be used for sitemaps, privacy practices, digital signatures, device characteristics, etc.

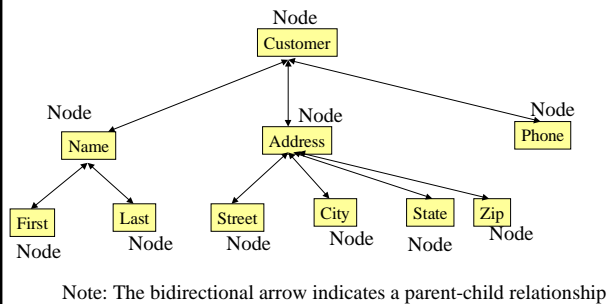
Copyright (A. Umar) 2002

Document Object Model

- DOM attaches behavior to Web documents
- Web documents viewed as objects
 - Elements
 - Attributes
 - Text
- OO API for accessing Web documents
 - DOM XML
 - DOM HTML DOM

Copyright (A. Umar) 2002

DOM



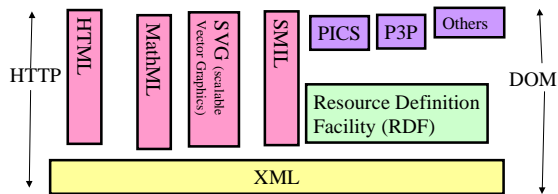
Copyright (A. Umar) 2002

HTTP Improvements

- HTTP 1.1 (with IETF) - Improve HTTP
 - Persistent connections
 - Support for virtual hosting (multiple DNS names to same IP address)
 - Sophisticated caching (control of what is cached how much)
- HTTP-NG (with Xerox)- Rethinking HTTP from scratch
 - Web as objects with IDL interfaces
 - Lightweight OO RPC
 - Several layers
- Other efforts: Web IDL, Web Brokers

Copyright (A. Umar) 2002

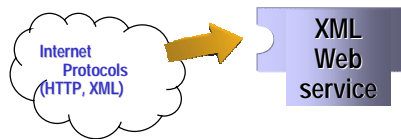
W3C Web Architecture



PICS: Platform for Internet Content Specification
 P3P: Platform for privacy preferences
 SMIL: Synchronous Multimedia Interaction Language
 DOM: Document Object Model

Copyright (A. Umar) 2002

XML Web Services



A programmable application component accessible via standard Web protocols

- Provide a directory of services on the Internet
- XML Web services are defined in terms of the formats and ordering of messages
- XML Web service consumers can send and receive messages using XML
- Built using open Internet protocols



Copyright (A. Umar) 2002

Standards Based

- SOAP (Simple Object Access Protocol)
 - Explicit serialization (HTTP + XML description) protocol used in service exchanges
- WSDL (Web Service Description Language)
 - XML document describing the location and interfaces a particular service supports – the client's contract
- DISCO (Discovery)
 - XML document describing (URI) of service
- UDDI (Universal Description Discovery and Integration)
 - Yellow pages directory for services

Copyright (A. Umar) 2002

Concluding Comments

- Internet and Web are foundation of Ecommerce
- Basic components of Web (HTML, HTTP, Browsers, Servers)
- Web is proceeding in several directions
- We have discussed two:
 - Improving human access
 - XML
 - Web automation
- Issue: will Web become too complex

References:

- www.w3.org
- IEEE Internet Computing, January 1999

Copyright (A. Umar) 2002
