1.264 Lecture 11

XML

What is XML?

- Extensible Markup Language (XML) is:
  - a World Wide Web Consortium (W3C) standard for
  - a file format to
  - easily and cheaply distribute electronic documents on the World Wide Web
  - extensible, not frozen like HTML
  - supporting rich structure, like objects or hierarchies or relationships
  - supporting validation and well-formed properties
  - avoiding applets, scripts, plug-ins, etc.
  - separating form (how it looks) from content (what it is)

Markup languages

- Many markup languages exist
  - MS Word uses Rich Text Format (RTF), a proprietary markup language (it now supports XML too)
  - WordPerfect
  - HTML
  - XML (Interleaf word processors have used XML for many years)
  - SGML
- These store tags (markup) in addition to raw text
- XML and HTML are subsets of SGML
  - XML provides 80% of SGML function with 20% of its complexity
  - SGML spec is 155 pages; XML spec is 35 pages
  - Removed all optional features of SGML

XML Concepts

- XML is self-describing and can be validated:
  - XML document contains the business rules to which its data must conformo
  - Rules can be reused in other documents: documents can be more specialized types (inheritance) of a base type
- XML applications
  - Data interchange format between computers
    - Using Web server as data channel between databases
    - Automated processing of documents exchanged
  - Common format for Web, electronic, paper documents, ...
    - XML as a general markup language
    - XML used for manuals, CDs, help and other text documents
    - Handled by standard browsers (IE, Mozilla, Netscape…)
  - Remote procedure call/invocation protocol
    - Executes Web services or processes on other computers

XML Toolso

- HTML displays data; XML describes data
- XML has associated languages
  - CSS, Cascading style sheets for display (pre-XSL)
  - XSLT: extensible stylesheet language/transformation
    - XML (without schemas) -> XML, HTML, or text
    - Scripting language, loosely typed, error tolerant
  - XQuery
    - XML (with schemas) -> XML
    - Strongly typed language, database access, ensures correctnesso
    - Major overlap with XPath (part of XSL), being coordinatedo
  - XML Schemao
    - Replaces DTDs (business rule representation), which are o not XML, with an XML definition
  - XHTML
    - Revised HTML that is valid XML
  - Others: RDF, XForms, … see www.w3c.org

XML Document Structure

- HTML:
    - <Head>
        - <Title>
    - <Body>

- XML (with DTDs—older style):
    - Prolog
        - XML declaration (defines version)
        - Document type declaration: DTD (defines tags)
    - Document body
        - Tags define data elements

XML Document Type Declarations (DTDs)

- Valid documents
  - Follows all rules:
    - e.g. #REQUIRED for an element
- Well formed document
  - Follows XML syntax but may not be valid
  - Used by browsers to accept XML documents that have already been validated by a server
    - No need to download the DTD and revalidate
  - Also used in small applications that can be written quickly
    - E.g. Departmental or workgroup document sharing

Example documents

- Product catalogs and availability
  - Product and service pricing and availability
- Order forms
  - Order data entry, invoicing, shipping, etc.
- Tax forms
  - Regulatory, customs, tax forms
- Books and data summaries
  - General technical data
- User manuals and instructions
  - Operating instructions for specific products and services
- Demo: 5 examples (2 email, 3 catalog) using XML
  - Use IE, which has the XML object model implemented
  - XSL not fully implemented and is ‘faked’ with scripts in demos

Dem

- Example 1: XML and HTML; XML and CSS
  - Change XML values; they change in HTML doc
  - Change CSS; it changes the appearance of HTML
- Example 2: XML and DTD
  - Entity (tag) names, cardinality (business rules)
    - Add illegal tag or fail to follow rule, get error!
- Example 3: XML and XML DSO (Microsoft) to traverse it
- Example 4: XML put in table with XML DSO
- Example 5: XML catalog as part of order entry
- Example 6: XML and XSL/T
- Example 7: Simple XSL

XML Document Example: Email

```
<?xml version="1.0"?>
<!DOCTYPE EMAIL SYSTEM "Example2.dtd">

<EMAIL LANGUAGE="Western" ENCRYPTED="128"
   PRIORITY="HIGH">
  <TO>Bharath@mit.edu</TO>
  <FROM>&SIGNATURE;@mit.edu</FROM>
  <CC>Darshan@mit.edu</CC>
  <BCC>gkocur@mit.edu</BCC>
  <SUBJECT>Sample Document with External DTD
   </SUBJECT>
  <BODY>
    Hello, this is &SIGNATURE;.
    Take care, -&SIGNATURE;
  </BODY>
</EMAIL>
```

Example: Catalog in XML

```
<parts>
  <part>
      <partNO>45891</partNO>
      <description>40mm Rubber Washer
  </description>          <inStock>Yes</inStock>
      <branch>Chicago, IL</branch>
      <cost>$0.45</cost>
  </part>
  <part>
      <partNO>40892</partNO>
      <description>45mm Rubber Washer
  </description>          <inStock>No</inStock>
      <branch<Crested Butte, CO</branch>
      <cost>$0.40</cost>
  </part>
</parts>
```

XML Implementation

- XML Engine
  - ActiveX or Java engine for Internet Explorer or Netscape
  - IE has XML installed by default
  - Netscape has implementation of XML, CSS, DSO, etc.
- XML Data Source Object (DSO) built in Microsoft IE
  - Connections to relational databases
  - Comparable capabilities in other browsers, clients
- XSLT
  - Includes query language to select data to display in browser
    - Select, comparison operators, order-by, …
- CSS
  - Includes format instructions to display XML as HTML
- Document Object Model (XML DOM)
  - Inheritance, node tree, …
- Database interface
  - Oracle, SQL Server, other databases handle XML directly

XML Schema

- XML Schema defines valid XML documents
- Supersedes DTDs, which are not XML documents themselves

XML file for XML Schema example

```
<Plant>
 <Wildflower>
  <Flower>
   Labrador tea
  </Flower>
  <Price>
   12.45
  </Price>
 </Wildflower>
 <Wildflower>
  <Flower>
   Dutchman's Breeches
  </Flower>
  <Price>
   3.89
  </Price>
 </Wildflower>
</Plant>
```

XML Schema example

```
<xsd:schema id="Plant"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:msdata="urn:schemas-microsoft-com:xml-msdata">
  <xsd:element name="Wildflower">
    <xsd:complexType>
      <xsd:all>
        <xsd:element name="Flower" minOccurs="1" type="xsd:string"/>
        <xsd:element name="Price" minOccurs="1" type="xsd:decimal"/>
      </xsd:all>
    </xsd:complexType>
  </xsd:element>
  <xsd:element name="Plant" msdata:IsDataSet="true">
    <xsd:complexType>
      <xsd:choice maxOccurs="unbounded">
        <xsd:element ref="Wildflower"/>
      </xsd:choice>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

.NET (ASPX) page to display XML

```
<%@ Import Namespace="System.Data" %>
<Script Runat="Server">
Sub Page_Load
  Dim dstMenu As DataSet
  dstMenu = New DataSet()
  dstMenu.ReadXmlSchema( MapPath( "PlantSchema.xml" ) )
  dstMenu.ReadXml( MapPath( "Plant.xml" ) )
  dgrdMenu.DataSource = dstMenu
  dgrdMenu.DataBind()
End Sub
</Script>

<html>
<head><title>PlantSchema.aspx</title></head>
<body>

<asp:DataGrid
  ID="dgrdMenu"
  cellpadding="10"
  Runat="Server" />
</body>
</html>
```

XML/EDI (Electronic Data Interchange)

- EDI is an old, mainframe-based standard for electronic exchange of information
  - EDI is expensive, complex and has limited usage, but enough that new technologies must interoperate with it
  - Very large companies use EDI and force it on their smaller partners/vendors
- XML documents follow EDI standards in many cases
- XML/EDI Maps XML <--> EDI and EDIFACT messages
- Mapping established using DTDs only
- Example:
  - EDItEUR book ordering messages
  - Bo k trade distribution, supply to libraries, new subscriptions, subscription renewals

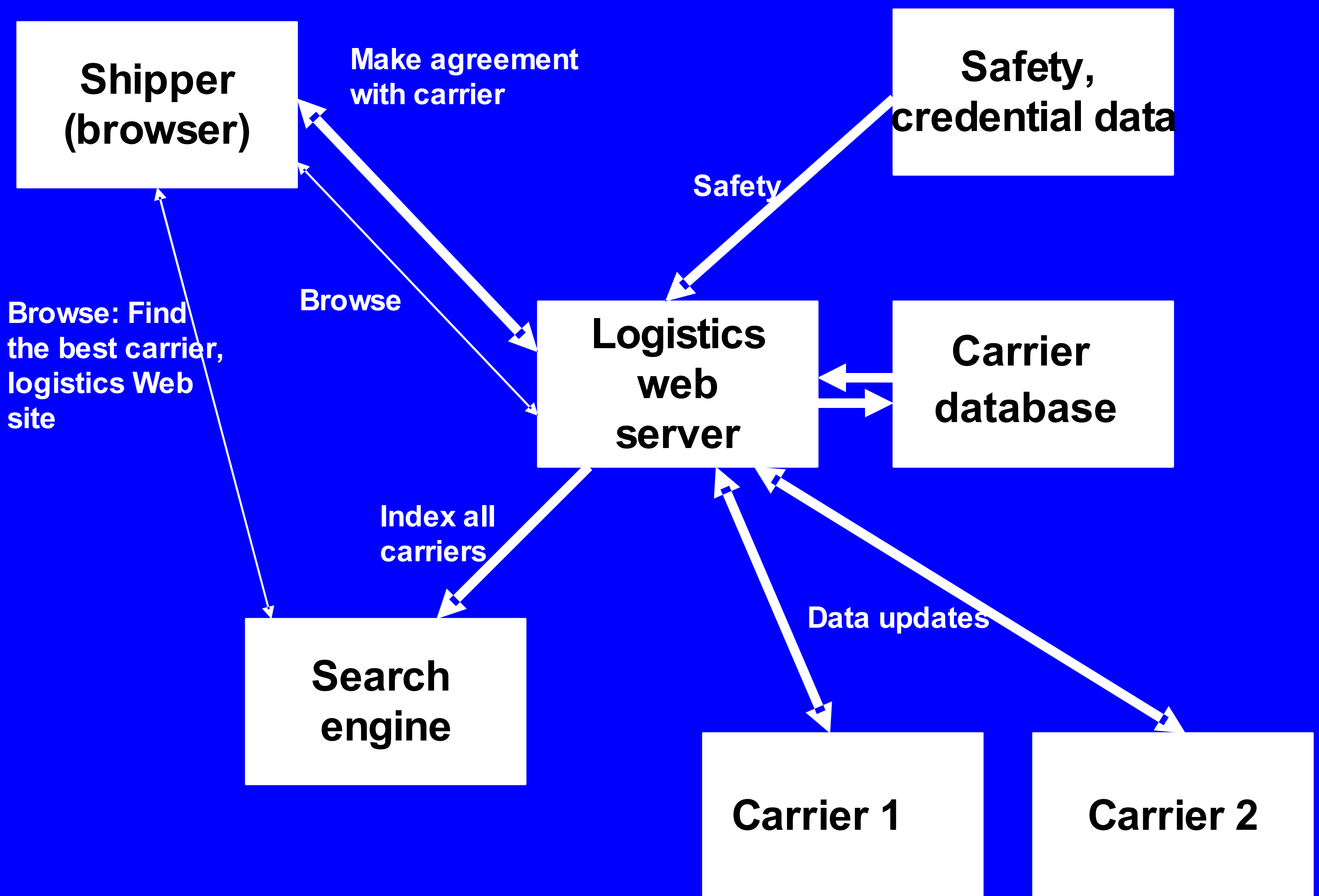
EDItEUR Message in XML

```
<!DOCTYPE Book-Order PUBLIC
   "-//EDItEUR//DTD Book Order Message//EN">
<Book-Order Supplier="8701"
   Send-to="http://www.mit.edu/order.jsp">
   <title>EDItEUR Lite-EDI Book Ordering </title>
   <Order-No> 934234 </Order-No>
   <Message-Date> 20000616 </Message-Date>
   <Buyer-EAN> 345345 </Buyer-EAN>
   <Order-line Reference-No="253535">
   <ISBN> 2352636464 6 </ISBN>
   <Author-Title> Smith, Steven/EDI Basics </Author-
   Title>
</Book-Order>
```

Supply chain management XML example

- End to end view of logistics
- Transportation stage composed of lanes
  - Service and volume packages put out to bid
  - Internet increasingly used as bidding mechanism
  - By making safety and credential data available via XML
    - Safety considerations are included in economic decisions, rather than after the fact
    - Carriers have incentives to have safety data, and to have good safety records
  - Government (e.g. safety and credential) data delivered via XML can tie naturally and easily into shipper and supply chain decision-making

Carrier selection in supply chain management



Trends

- With XML, Internet will support distributed document processing as its standard information exchange mechanism
- Interbusiness ecommerce will be biggest user of this capability
  - Documents are a natural and current means of interbusiness communications
- New users of Internet ecommerce may well skip EDI and use cheaper XML
  - Less than 80,000 of 6.2 million US businesses use EDI
  - 125,000 businesses world-wide use EDI
  - EDI cost and complexity are insurmountable obstacle for small and medium size businesses

Protocols for next generation Internet business

- HTTP, moving to SOAP (next lecture)
- HTML, moving to XHTML, for display
- XML
  – XML import and export to/from databases
  – Mapping XML to legacy formats and back
  – Mapping between different XML DTDs and XSchema
  – DTDs and tools for specific ecommerce areas
  – XSLT query language to select XML data to display (SQL-like) and for display itself
- Security protocols (later)