What is XML?

- XML – eXtensible Markup Language
- Definition: XML is a markup language for documents containing structured information
- Markup = "tags" embedded in text content of a document to its structure
- A subset of the Standard Generalized Markup Language (SGML), which is a ISO standard and widely used in the publishing industry
- XML is a W3C standard. Latest version is 1.1
  - http://www.w3.org/TR/2006/REC-xml11-20060816/
Why XML?

- HTML is the standard for web browsers, why use another markup language?
  - HTML is used primarily for presentation purposes, e.g. for human-machine interactions
  - HTML can not support machine-to-machine (M2M) data interactions effectively
  - Document structure can not be effectively represented using HTML

- What is a document and why its structural representation is so important?
  - Document = Structured Information = Computer File
  - To support M2M data exchange, the structure of the document has to be clearly defined

Sample Doc Structure (JFIF)

<table>
<thead>
<tr>
<th>Field</th>
<th>Size (bytes)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>APP0 Marker</td>
<td>2</td>
<td>Always equals 0xFFE0</td>
</tr>
<tr>
<td>Length</td>
<td>2</td>
<td>Length of segment excl. APP0</td>
</tr>
<tr>
<td>Identifier</td>
<td>5</td>
<td>Always equals &quot;JFIF&quot; in Hex</td>
</tr>
<tr>
<td>Version</td>
<td>2</td>
<td>1st byte is major, 2nd byte is minor</td>
</tr>
<tr>
<td>Density Units</td>
<td>1</td>
<td>Units for pixel density fields</td>
</tr>
<tr>
<td>X Density</td>
<td>2</td>
<td>Integer horizontal pixel density</td>
</tr>
<tr>
<td>Y Density</td>
<td>2</td>
<td>Integer vertical pixel density</td>
</tr>
<tr>
<td>Thumbnail Width</td>
<td>1</td>
<td>Horizontal size of thumbnail in pixels</td>
</tr>
<tr>
<td>Thumbnail Height</td>
<td>1</td>
<td>Vertical size thumbnail in pixels</td>
</tr>
<tr>
<td>Thumbnail Data</td>
<td>3 * T-Width * T-Height</td>
<td>Uncompressed 24 bit RGB raster thumbnail</td>
</tr>
</tbody>
</table>
EDI X12 820 (Payment Order)

- ISA*00* 000* 01*123454321*01*012341234
  0*031016*2359*U*00401*987600111*X*
  0*04010
  IST*820*987600111
  BPR*G*77.77*C*AGHCTX*01*234056789*DA*0099109999*
  1*123453421*01*045678099*DA*1008973899*031016
  TRN*1*0310162359
  REF*AA*ED16
  N1*PR*WHIZCO OF AMERICA INC
  N3*55 MEGAPLEASANT ROAD*SUITE 999
  N4*SUPERVILLE*NY*10954
  N1*PE*YOWZACO
  MENT*1
  RMR*AP*111111111111111111PO*11.11
  RMR*AP*222222222222222222PO*22.22
  RMR*AP*444444444444444444PO*44.44
  DTM*055*031016
  SE*000000014*987600111
  GE*1*987600111
  IEA*1*987600111

**Total PO amount = $77.77**

- Address and location info

**Itemized POs**

- ISA Interchange Control Header
- GS Function Group Header
- ST Transaction Set Header
- SE Transaction Set Trailer
- GE Function Group Trailer
- IEA Interchange Control Trailer

= Segment Delimiter
* = Field Delimiter

**Steps in M2M Data Exchange**

1. **Input Stream**
2. **Pre-Processing**
   - Parser
   - Internal Data Representation of Document Objects
3. **Post Processing**
   - Data Processor

- The generic model for document processing
- Modularity/Reuse is achievable via standardization
Document Pre-processing

- Post-processing is application-specific (limited reusability)
- An EDI parser has to be constructed manually for different types of EDI document, because of its structural complexity, even EDI is standardized
- Even for minor structural changes (e.g. adding a field in a segment), the parser has to be reconstructed
- The structure of EDI document is defined on paper in non-electronic format

A document framework for preprocessing

- A generic framework is needed in support of M2M data exchange
- Standardization of basic encoding format
  - Text-based encoding for document representation, since it is more flexible and platform/application neutral than binary format
- Representation of structural information within the document - use predefined tags in docs
- Codify the structural information of documents
  - Refer as the document metadata
  - Enable the dynamic construction of doc parser
XML-based document framework

XML Input File

Document Object Model (DOM)

XML Parser

Data Processor

- Document Type for XML – XML Schema or Document Type Definition (DTD)

Type definition using XML Schema

```xml
<?xml version="1.0" encoding="UTF-8"?>
<PurchaseOrder xsi:noNamespaceSchemaLocation="PurchaseOrder.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <RequisitionNo>485321</RequisitionNo>
  <Product>
    <SkuNo>U2322</SkuNo>
    ...
  </Product>
</PurchaseOrder>
```

The default namespace is represented by the schema file

xs/ refers to the schema instance defined by the XML schema standard

Type definition using DTD

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE PurchaseOrder SYSTEM "PurchaseOrder.dtd">
<PurchaseOrder>
  <RequisitionNo>485321</RequisitionNo>
  <Product>
    <SkuNo>U2322</SkuNo>
    ...
  </Product>
</PurchaseOrder>
```

SYSTEM refers to the local DTD definition

**An XML Example**
What is XML Schema?

- An abstract data model consists of the basic building blocks for specifying document type definitions of XML documents
- Enable document validation, making XML a “Strongly typed” language
  - Schema files are in XML format with the ext. xsd
- Designed to be scalable and extensible in an object-oriented fashion
  - xsi refers to the basic schema for all schema defs
- More widely used than DTD since it is designed for M2M data exchange
Pre-Processing of XML Document

- Document validation can be enforced by stream parser
- Data binding refers to mapping XML elements to language-specific data structure, e.g. mapping XML elements in a document to Java objects
- JAXB – Java XML Binding

Applications using XML

- M2M data exchange
  - Web Services, RSS
  - AJAX, Data Mashups (e.g. GoogleMaps)
- Document processing and presentation
  - Document translation via XSLT
  - Dynamic publication of document in HTML
- Software Development and System Admin
  - Declarative programming
  - Software component configuration
  - Software process control (e.g. Ant, Maven)
  - System and Security/access control configuration
Data and Service Mashups

- What is mashup? A web application that combines data from more than one source into a single integrated tool (Wikipedia)
- In general, mashup is the concept of integrating data from multiple data sources to create presentation in a new dimension
  - Data mashup for presentation
  - Service mashup for reuse
- Client-side mashup: AJAX, GoogleMaps, AdobeFlash, etc.
- Server-side mashup: Proxy or gateway services providing data mapping, conversion, and mediation functions

Client-side Mashups

- Information from multiple sources is consolidated on user’s web browser
Google Maps APIs

- A set of JavaScript APIs using AJAX concept (based on XMLHttpRequest)
- To embed Google Maps on your own web site using JavaScript
- To overlay geospatial information from any web source in XML format

Sample Google Maps Script

```html
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xmlns:v="urn:schemas-microsoft-com:vml">
  <head>
    <meta http-equiv="content-type" content="text/html; charset=utf-8"/>
    <title>Google Maps JavaScript API Example: Simple Map</title>
    <script src="http://maps.google.com/maps?file=api&amp;amp.v=2&amp;amp.key=ABQIAAAAzr2EBOXUKnm_jVnk00JfXxSosDVG8kkPE1-m51RBnYughuyMxQ-H1QfUnh94QxWltbN4U6MouMmBA" type="text/javascript"></script>
    <script type="text/javascript">
      function initialize() {
        if (GBrowserIsCompatible()) {
          var map = new GMap2(document.getElementById("map_canvas"));
          map.setCenter(new GLatLng(37.4419, -122.1419), 13);
        }
      }
    </script>
  </head>
  <body onload="initialize()" onunload="GUnload()">
    <div id="map_canvas" style="width: 500px; height: 300px"></div>
  </body>
</html>
```
Overlay XML Data onto Map

function initialize() {
    if (GBrowserIsCompatible()) {
        var map = new GMap2(document.getElementById("map_canvas"));
        map.setCenter(new GLatLng(37.4419, -122.1419), 13);
        map.addControl(new GSmallMapControl());
        map.addControl(new GMapTypeControl());

        // Download the data in data.xml and load it on the map. The format we
        // expect is:
        // <markers>
        //   <marker lat="37.441" lng="-122.141"/>
        //   <marker lat="37.322" lng="-121.213"/>
        // </markers>
        GDownloadUrl("./include/data.xml", function(data) { 
            var xml = GXml.parse(data);
            var markers = xml.documentElement.getElementsByTagName("marker");
            for (var i = 0; i < markers.length; i++) {
                var latlng = new GLatLng(parseFloat(markers[i].getAttribute("lat")),
                                            parseFloat(markers[i].getAttribute("lng")));
                map.addOverlay(new GMarker(latlng));
            }
        });
    }
}

Online References

- W3C XML (1.1) Specification, 2nd edition
  - [http://www.w3.org/TR/2006/REC-xml11-20060816/](http://www.w3.org/TR/2006/REC-xml11-20060816/)
- Introduction of XML
  - [http://www.xml.com/pub/a/98/10/guide0.html?page=1](http://www.xml.com/pub/a/98/10/guide0.html?page=1)
- XML Schema
  - Primer [http://www.w3.org/TR/xmlschema-0/](http://www.w3.org/TR/xmlschema-0/)
  - Tutorial [http://www.w3schools.com/schema/default.asp](http://www.w3schools.com/schema/default.asp)
- GoogleMaps API and Developer’s Guide
Summary

- XML provides a standard way of presenting structured document for M2M data exchange
- XML Schema defines the type information of XML document in XML (codified format). It is designed in an extensible, object-oriented fashion, and enables XML to be validated and become "strongly typed"
- The codification of document types enables dynamic configuration of document parsers and dynamic data binding in document pre-processing
- XML technology enables data mashups from different distributed data sources (based on XMLHttpRequest)
- GoogleMaps APIs provide a rich set of JavaScript tools for mashing up geospatial information with 2D maps, creating an eco environment for data sharing

Quiz

- What is the standard for representing an XML document in data structures?
- What are the steps necessary to create a DOM structure of an XML document in computer memory?
- What is the GoogleMaps API for retrieving a set of geo-referenced data in XML format?
- Create a GoogleMaps web page with an overlay of hospital information in the New Orleans region (XML file on class web site)