Part C: Introduction to XML

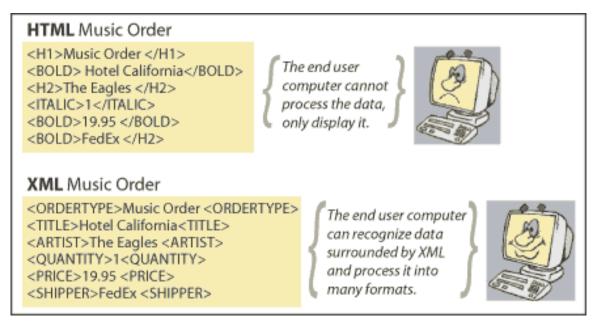
I. Weakness of HTML

- HTML is not a suitable language for making data meaningful to computer programs.
- This is a serious shortcoming because the whole business world (banking, insurance, retail, etc) is dependent on computer programs interpreting data.

II. Definition of XML

XML stands for eXtensible Markup Language. XML was designed for carrying and storing data. However, HTML was designed to display data. XML has following features:

- XML is a markup language much like HTML.
- XML tags are not predefined. You must define your own tags.
- XML is designed to be self-descriptive.



(Ref: www.information-management.com)

III. Advantages of XML

- **Data awareness:** since XML is self describing it is possible for programs that process them to act more "intelligently".
- Independence of communicating parties: XML is independent of all machines, operating systems, programming languages and databases.

• **Standard language:** XML is being used to define standard languages, vocabularies, for sharing data by many industry sectors and professional groups (e.g., Mathematical Markup Language (MathML), Open Financial Exchange (OFX), etc).

IV. XML Document Structure

- XML documents form a tree structure that starts at "the root" and branches to "the leaves".
- An Example XML Document:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

<note>

<to>Mary</to>

<from>Tom</from>

<heading>Reminder</heading>

<body>Don't forget the meeting on this Friday!</body>

</note>

- * 1st line: the XML declaration. It defines the XML version (1.0) and the encoding used (ISO-8859-1 = Latin-1/West European character set).
- * 2nd line: it describes the root element of the document (like saying: "this document is a note").

- * The next 4 lines describe 4 child elements of the root (to, from, heading, and body).
- * The last line defines the end of the root element:

V. XML Syntax Rules

- XML Documents Must Have a Root Element: XML documents must contain one element that is the parent of all other elements. This element is called the root element.
- All XML Elements Must Have a Closing Tag.
- XML Tags are Case Sensitive: XML tags are case sensitive. With XML, the tag <a href="Lett
- XML Elements Must be Properly Nested.

Wrong: <i>This text is bold and italic</i>

Correct: <i>This text is bold and italic</i>

• XML Attribute Values Must Be Quoted: In XML the attribute value must always be quoted.

Wrong: <note date=12/11/2007>

Correct: <note date="12/11/2007">

• Entity References: Some characters have a special meaning in XML.

Wrong: <message>if salary < 1000 then</message>

Correct: <message>if salary < 1000 then</message>

Entity Reference	Symbol	Meaning
<	\	less than
>	>	greater than
&	&	ampersand
'	•	apostrophe
"	"	quotation mark

- Comments in XML: The syntax for writing comments in XML is similar to that of HTML. Example: <!-- This is a comment -->
- With XML, White Space Is Preserved: HTML reduces multiple white space characters to a single white space:
- XML Stores New Line as LF (line feed): in Windows applications, a new line is normally stored as a pair of characters: carriage return (CR) and line feed (LF).

VI. XML Elements

- An XML element is everything from (including) the element's start tag to (including) the element's end tag.
- An element can contain other elements, simple text or a mixture of both. Elements can also have attributes.
- Naming Rules for Elements
 - * Names can contain letters, numbers, and other characters
 - * Names cannot start with a number or punctuation character
 - * Names cannot start with the letters xml (or XML, or Xml, etc)
 - * Names cannot contain spaces

VII. XML Attributes

- XML elements can have attributes in the start tag. Attributes provide additional information about elements.
- Attributes are difficult to read and maintain. It is advised that we should use elements for data, and use attributes for information that is not relevant to the data.
- Elements vs. Attributes (demonstrate with examples)

```
Example with attributes:
<person sex="female">
<firstname>Anna</firstname>
<lastname>Smith</lastname>
</person>

Example with elements:
<person>
<sex>female</sex>
<firstname>Anna</firstname>
<lastname>Smith</lastname>
</person>
```

VIII. XML Validation

- A "Valid" XML document is a "Well Formed" XML document, which also conforms to the rules of a Document Type Definition (DTD).
- The purpose of a DTD is to define the structure of an XML document. It defines the structure with a list of legal elements.
- Example:

The DOCTYPE declaration in the example above is a reference to an external DTD file. The content of the file is shown in the paragraph below (where PCDATA = Parsed Character DATA).

```
<!DOCTYPE note [
  <!ELEMENT note (to,from,heading,body)>
  <!ELEMENT to (#PCDATA)>
  <!ELEMENT from (#PCDATA)>
  <!ELEMENT heading (#PCDATA)>
```

```
<!ELEMENT body (#PCDATA)>
]>
```

IX. Viewing XML Files

- A XML file should have a file extension of ".xml".
- Raw XML files can be viewed in all major browsers (such as Firefox, IE, etc).