Introduction to XML

What is XML?

- XML stands for eXtensible Markup Language
- XML is a simple, very flexible text format
- XML is actually a metalanguage – a language for describing other languages
- XML is a text-based markup language
- Unlike HTML, XML tags identify the data, rather than specifying how to display it

What is XML?

- eXtensible – It is extensible because it is not a fixed format like HTML
- Markup - A markup language is a set of words and symbols for describing the identity of pieces of a document (for example ‘this is a paragraph’, ‘this is a heading’, etc).

Why XML?

- Plain Text - being a text format, XML is both computer and human readable
- Stylability - XSL lets you dictate how to display XML data
- Easily Processed - regular and consistent notation makes it easier to build a program to process XML data

An example XML Document

```
<note>
    <to>Tove</to>
    <from>Jani</from>
    <heading>Reminder</heading>
    <body>
        Don't forget me this weekend!
    </body>
</note>
```

An example XML Document

```
<note>
      <to>Tove</to>
      <from>Jani</from>
      <heading>Reminder</heading>
      <body>
          Don't forget me this weekend!
      </body>
</note>
```

The first line describes the root element of the document (like it was saying: "this document is a note").

An example XML Document

```
<note>
    <to>Tove</to>
    <from>Jani</from>
    <heading>Reminder</heading>
    <body>
        Don't forget me this weekend!
    </body>
</note>
```

The next 4 lines describe 4 child elements of the root (to, from, heading, and body).

An example XML Document

```
<note>
      <to>Tove</to>
      <from>Jani</from>
      <heading>Reminder</heading>
      <body>
          Don't forget me this weekend!
      </body>
</note>
```

In XML, every tag must have a closing tag.

An example XML Document

```
<note>
      <to>Tove</to>
      <from>Jani</from>
      <heading>Reminder</heading>
      <body>
          Don't forget me this weekend!
      </body>
</note>
```

Note, too, that the content of the <to> tag is entirely contained within the scope of the <note>…</note>

An example XML Document

```
<note>
      <to>Tove</note>
      <from>Jani</from>
      <heading>Reminder</heading>
      <body>
          Don't forget me this weekend!
      </body>
</to>
```

This is an error!!!

An example XML Document

```
<note date="24/04/2006">
      <to>Tove</to>
      <from>Jani</from>
      <heading>Reminder</heading>
      <body>
          Don't forget me this weekend!
      </body>
</note>
```

XML tags can have attributes to better describe data.

An example XML Document

```
<note to="Tove" from="Jani"
heading="Reminder">
      <body>
          Don't forget me this weekend!
      </body>
</note>
```

You could design a data structure like <note> equally well using either attributes or tags.

An example XML Document

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<note date=”24/04/2006”>
    <to>Tove</to>
    <from>Jani</from>
    <heading>Reminder</heading>
    <body>
        Don't forget me this weekend!
    </body>
</note>
```

An XML file always starts with a prolog. The minimal prolog is:
<?xml version="1.0"?>

An example XML Document

```
<?xml version="1.0" encoding="ISO-8859-1">
<note date=”24/04/2006”>
    <to>Tove</to>
    <from>Jani</from>
    <heading>Reminder</heading>
    <body>
        Don't forget me this weekend!
    </body>
</note>
```

Identifies the version of the XML markup language used in the data. This attribute is not optional.

An example XML Document

```
<?xml version="1.0" encoding="ISO-8859-1">
<note date=”24/04/2006”>
    <to>Tove</to>
    <from>Jani</from>
    <heading>Reminder</heading>
    <body>
        Don't forget me this weekend!
    </body>
</note>
```

(Optional) Identifies the character set used to encode the data. "ISO-8859-1" is "Latin-1" the Western European and English language character set (the default is compressed Unicode: UTF-8).

XML Validation

- A "Well Formed" XML document has correct XML syntax.
- A "Valid" XML document also conforms to a DTD (Document Type Definition)

DTD

- With DTD, each of your XML files can carry a description of its own format with it.
- With a DTD, independent groups of people can agree to use a common DTD for interchanging data.
- Your application can use a standard DTD to verify that the data you receive from the outside world is valid.
- You can also use a DTD to verify your own data.

DTD

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE note SYSTEM "note.dtd">
<note date=”24/04/2006”>
    <to>Tove</to>
    <from>Jani</from>
    <heading>Reminder</heading>
    <body>
        Don't forget me this weekend!
    </body>
</note>
```

Where should I use XML?

- Information identification
- Information storage
- Information structure
- Publishing
- Messaging and data transfer
- Web services

References

W3C – http://www.w3.org/TR/REC-xml/

FAQ – http://xml.silmaril.ie/

Thank you :-)