XML History

XML is the Extensible Markup Language.

HTML was derived from SGML, the Standard Generalized Markup Language.

XML was also derived from SGML.

Version 1.0 of XML was released in Feburary 1998.

Some related technologies:

XSL	Extensible Stylesheet Language
XSLT	XSL Transformations
XSL-FO	XSL Formatting Objects
SAX	Simple API for XML

XML Basics

Here is a simple XML *document*:

```
<language>
Icon
</language>
```

The document has one *element*. The name of the element is `language`.

`<language>` is a *start-tag* and `</language>` is an *end-tag*.

Every element must have a start-tag and an end-tag.

The *content* of an element is everything between the start-tag and the end-tag of the element.

The content of the `language` element is the *character data* `Icon`.

Element names can contain "letters", numbers, underscores, hyphens (dashes), and periods. Names must begin with an underscore or a letter.

Letters need not be in English:

```
<π>
Apple
</π>
```

XML Basics, continued

Here is another XML document:

```
<colors>
    <color>Red</color>
    <color>Green</color>
    <color>Blue</color>
</colors>
```

The *root element* of this document is named `colors`. Every XML document has one and only one root element.

The element `colors` contains three elements, each named `color`.

Each `color` element contains character data with a color name.

No whitespace may appear in a start-tag or an end-tag.

XML Basics, continued

An element may have *attributes* associated with it.

```
<rectangle color="red" filled = 'no' >
<WidthHeight>10 20</WidthHeight>
</rectangle>
```

The element `rectangle` has two attributes: `color` and `filled`. Attribute names follow the same rules as element names.

An attribute name is followed by an equals sign, then a value in quotes. Either single or double quotes may be used. Whitespace may surround the equals sign.

There are no rules restricting when data should be present as attributes versus content. One might choose this representation instead:

```
<rectangle color="red" filled = 'no'
    w="10" h="20">
</rectangle>
```

Or this:

```
<rectangle>
  <color>red</color>
  <filled value= 'no'></filled>
  <Width>10</Width>
  <Height>20</Height>
</rectangle>
```

Or this:

```
<rectangle>red no 10 20</rectangle>
```

XML Basics, continued

Empty elements have no content. An *empty-element tag* starts with "<" and ends with "/>". Here is a complete XML document that consists of one empty element:

```
<X/>
```

This is equivalent:

```
<X></X>
```

Empty elements may have attributes:

```
<rectangle color=’red’ w=’10’ h=’20’ />
```


XML Basics, continued

The content of an element may be an arbitrary mix of elements and character data:

```
<x>abc
    <d/>
    efg
    <d></d>
    <f>
        <abc>x</abc>
    </f>
    xyz
    PDQ
</x>
```

Here is an analysis of the contents:

```
Element: x
    Char Data: "abc"
    Element: d
    Char Data: "efg"
    Element: d
    Element: f
        Element: abc
            Char Data: "x"
    Char Data: "xyz\n    PDQ"
```

Technically speaking, whitespace is included in the character data, but the above analysis has taken some liberties with it.

XML Basics, continued

Comments may appear in XML. Here is a comment:

```
<!-- This is a comment -->
```

Comments may span many lines. The string "--" may not appear in a comment.

Comments may appear anywhere that character data may appear.

Comments may also appear before and after the root element.

Comments may not appear inside a tag.

Example:

```
<!-- Old colors....
<colors>
    <color>Red</color>
    <color>Green</color>
    <color>Blue</color>
</colors>
-->
<colors>
    <color>Black</color>
    <color>White</color>
    <!-- Who needs more colors?? -->
</colors>
```


XML Basics, continued

NOTE:

This was a very thin introduction to XML. There is much more to it.