

A Brief Introduction to XML

Bruce M. Bolden

January 14, 2010

1 Introduction

XML [[XML 98](#)], [[Walsh 98](#)] is rapidly becoming a standard format for storing information. Apple currently uses XML to store property lists (plist) to hold the preferences for applications and the OS X operating system. Microsoft is using it in .Net (their next generation framework). A number of Open Source applications use configuration files defined using XML.

Originally, XML was designed as a markup language for text documents that contained structured information (the *X* in *XML* stands for extensible). Data is included in XML documents as strings of text and the data is surrounded by markup text that describes the data. A particular unit of data and markup is called an *element*.

XML is easily transformed into other forms using XSLT (Extensible Stylesheet Language Transformations) [[GW00](#)]. XSLT will not be discussed further in this document.

Applications that try to understand the contents of an XML document typically use an XML parser to read the document. The parser break up the document into individual elements. A Document Type Definition (DTD) specifies the format of the data. The DTD or a link to a DTD is normally included in the XML document.

2 XML Basics

An XML document contains text, never binary data. XML is case sensitive. A very simple, complete XML document:

```
<course>
CS 121---Data Structures
</course>
```

2.1 Tag Syntax

XML tags look like HTML tags. Start tags begin with a `<` and end tags begin with a `</`. The start tag symbol is followed by the element name. The tag is finished with a `>`.

2.2 XML Trees

XML documents are trees. This is illustrated in the following document:

```
<person>
  <name>
    <first_name>Bruce</first_name>
    <last_name>Bolden</last_name>
  </name>
  <profession>Computer Scientist</profession>
  <profession>Teacher</profession>
</person>
```

This XML document is still composed of one `person` element, but now contains three *child* elements: a `name` and two `profession` elements.

2.2.1 Example

A simple “Hello World” defined in XML for use with Glade (an interface builder for the open-source GTK toolkit library for X) [[ESR 04](#)].

```
<?xml version="1.0"?>
<GTK-Interface>

<widget>
  <class>GtkWindow</class>
  <name>HelloWindow</name>
  <border_width>5</border_width>
  <Signal>
```

```

    <name>destroy</name>
    <handler>gtk_main_quit</handler>
</Signal>
<title>Hello</title>
<type>GTK_WINDOW_TOPLEVEL</type>
<position>GTK_WIN_POS_NONE</position>
<allow_shrink>True</allow_shrink>
<allow_grow>True</allow_grow>
<auto_shrink>False</auto_shrink>

<widget>
  <class>GtkButton</class>
  <name>Hello World</name>
  <can_focus>True</can_focus>
  <Signal>
    <name>clicked</name>
    <handler>gtk_widget_destroy</handler>
    <object>HelloWindow</object>
  </Signal>
  <label>Hello World</label>
</widget>
</widget>

</GTK-Interface>

```

3 References

References

- [XML 98] XML.com, <http://www.xml.com/>
- [XML 10] Extensible Markup Language (XML) 1.0 (Fifth Edition)
<http://www.w3.org/TR/REC-xml/>
- [Walsh 98] *A Technical Introduction to XML*, Norman Walsh,
<http://www.xml.com/pub/a/98/10/guide0.html>

- [ESR 04] The Art of Unix Programming, Eric S. Raymond,
<http://www.catb.org/~esr/writings/taoup/>
- [GW00] XSL Concepts and Practical Use, Paul Grosso and Norman
Walsh,
<http://nwalsh.com/docs/tutorials/xsl/xsl/slides.html>