

XML 1.0 Quick Reference

Well Formed Versus Valid

Must be well formed. Can also be valid with respect to a validating document, e.g. an XSD schema.

Start

Declaration

```
<?xml version="1.0" encoding="UTF-8"
standalone="no"?>
```

Should have.

Character encoding of the document, expressed in Latin characters, e.g. utf-8, utf-16, iso-8859-1

Standalone declaration:
no: parsing affected by external DTD subset
yes: parsing not affected by external DTD subset

Processing Instruction

```
<?xml-stYLESHEET type="type" href="url" title="string"?>
```

```
<?xml-stylesheet href="/style.css" type="text/css"
title="default stylesheet"?>
<?xml-stylesheet href="/style.xml" type="text/xml"
title="default stylesheet"?>
```

Namespaces

Declaring

```
<element xmlns="uri" xmlns:prefix="uri"> ... </element>
```

```
<!-- No namespace -->
<bar
```

```
<!-- Default namespace -->
<bar xmlns="http://bar"
```

```
<!-- Explicit namespace -->
<bar xmlns:ba="http://bar"
```

Default namespace

There can only be one default namespace.

Elements and Attributes without a prefix

Elements without a prefix will either belong to the:
* Default namespace; or

* No namespace when no namespace is specified in a containing element.

Attributes without a prefix will belong to no namespace.

Referencing an Explicit Namespace

Use declared prefix before the element or attribute.

```
<cloud xmlns:wx="http://example.org/wx">
  <wx:height wx:units="feet">2500</wx:height>
</cloud>
```

Namespace Name (URI)

A namespace name is a Universal Resource Identifier which is either a:

1. A uniform resource name (URN)
<URN> ::= "urn:" <NID> ":" <NSS>

where <NID> is the Namespace Identifier, and <NSS> is the Namespace Specific String.

```
urn:www-develop-com:student
urn:www.ed.gov:elementary.students
urn:uuid:E7F73B13-05FE-44ec-81CE-F898C4A6CDB4
```

Or

2. A Permanent uniform resource location (PURL)
http://www.develop.com/student
http://www.ed.gov/elementary/students

XSD Schema References

```
<rootelement
  xmlns:[prefix]="uri-reference"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
  xsi:schemaLocation="uri-reference schemaLocation">
```

```
<!-- No namespace -->
<bar
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
  xsi:noNamespaceSchemaLocation="bar.xsd">
```

```
<!-- Default namespace -->
<bar xmlns="http://bar"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
  xsi:schemaLocation="http://bar bar.xsd">
```

```
<!-- Explicit namespace -->
<bar xmlns:ba="http://bar"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
  xsi:schemaLocation="http://bar bar.xsd">
```

Body Rules

The rest of this quick reference describes the rules a document needs to conform to in order to be well formed.

Comment

```
<!-- Stuff here -->
```

Elements

Root Element

There must be one and only one root element.

No Overlapping Elements

Wrong:
<galaxy>Milky way <planet> earth
</galaxy></planet>

Right:
<galaxy>Milky way
<planet> earth</planet>
</galaxy>

Elements must have a start and end tag or...

```
<element> ... </element>
Wrong:
<p> This is an HTML paragraph.
<p>or two.
Right:
<p>This is an HTML paragraph.</p>
<p>or two.</p>
```

... Be an empty element

```
<element />
```

For XHTML backward compatibility include a space before the slash.

Mixed Content

Elements may have mixed content. That is, text not surrounded by any child:
<galaxy>**Milky way**
<planet>earth</planet>
</galaxy>

Casing

Element and attribute names are case-sensitive.

Attributes

Attribute values must always be quoted. Either ' or ". (I prefer ").
Wrong:

Right:

If an attribute exists it must have a value (this can be an empty string).

Wrong:
<aircraft engines />

Right:
<aircraft engines="4" />
<aircraft engines="" />

CDATA Section

```
<![CDATA[ ... ]]>
```

A CDATA section designates that the text within it should be not marked up. Common uses for CDATA sections are for embedding: script; styles; and XML or XHTML examples.

Character References

There are three ways to refer to a character: literally; a Numeric Character Reference and a Named Entity Reference.

Numeric Character Reference (with a numeric code):
€ ' € - Decimal Form &#value;
€ ' € - Hexadecimal Form &#xvalue;

 ' Line Feed (a new line in XML)

The five predefined Named Entity References (with an alphabetic code) are :

```
&lt; ' <
&gt; ' >
&amp; ' &
&quot; ' "
&apos; ' '
```

' and ' can be referred to literally in element content.

Unicode

Xml works in unicode.

Whitespace

White space is not allowed before the XML declaration.

In element content white space can be designated as requiring preservation or that preservation is up to the default behaviour of the XML parser.

```
<element xml:space="preserve|default"> ...
</element>
```

In attributes XML processors will frequently report Tabs, carriage returns, and spaces as single spaces.

The two character sequence Carriage Return-Line Feed (CRLF) and the single character sequences CR or LF will all be reported as a single LF character.

Example

```
<?xml version="1.0" encoding="utf-8"
standalone="no"?>
<?xml-stylesheet href="/style.css" type="text/css"
title="default stylesheet"?>
<?xml-stylesheet href="/style.xml" type="text/xsl"
title="default stylesheet"?>
<!DOCTYPE catalog SYSTEM "catalog.dtd">
<!-- This is a comment -->
<!-- Document Elements -->
<catalog
xmlns:myns="urn:www-me-com:myeffort"
xmlns="http://www.example.com/catalog/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
xsi:schemaLocation="
http://www.example.com/catalog/
http://tempuri.org/catalog.xsd">
  <book id="bk101">
    <author>&#71;Abercrombie, Kim</author>
    <title>XML Developer's &#x47;uide</title>
    <genre>Computer</genre>
    <price>44.95</price>
    <publishDate>2000-10-01</publishDate>
    <description><![CDATA[An in-depth look
      at creating application
      with XML, using <, >,
      and &amp;.]]>
    </description>
  </book>
  <book id="bk109">
    <genre>Science Fiction</genre>
    <myns:price>6.95</myns:price>
    <publishDate>2000-11-02</publishDate>
    <description>This is a great read and a fountain
      of knowledge.
    </description>
  </book>
</catalog>
```