

How do I format XML?

The Rules



■ XML Document Structure

All XML documents must be “Well-formed”

The five (5) basic XML Document Rules:

1. Tags are case-sensitive
2. Opening tags must have closing tags
3. Tags must be properly nested
4. Attribute values must be quoted
5. Root “document” element must be *unique*, and is *required*.

Rule #1 -

- Tags are case sensitive

Dog is different from dog which is different from DOG

```
<Message>This is incorrect</message>
```

```
<message>This is correct</message>
```



Rule #2

Opening tags must have a closing tag

```
<firstName>Bruce  
<lastName>Wayne
```

```
<firstName>Bruce</firstName>  
<lastName>Wayne</lastName>
```



Rule #2

- You can include whitespace after the tag value (both opening and closing)

`<first >Clark</first >`

OR

`<first`

`>Clark</first>`



- Spaces are NOT allowed preceding the tag value

`< first>Clark< /first>`



Rule #3

- Tags must be properly nested

If you start element A, then start element B, you must first close element B before closing element A.

```
<book><title>This is not properly nested</book></title>
```

```
<book><title>This is properly nested.</title></book>
```



Rule #4

- Attribute values must be quoted

```
<note date=12/11/2007>  
  <to>Dave</to>  
  <from>John</from>  
  </note>
```

```
<note date="12/11/2007">  
  <to>Dave</to>  
  <from>John</from>  
  </note>
```

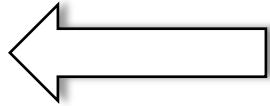


They can be single OR double – as long as they match each other

Rule #5

- Root element required and it must be unique – the beginning of your XML document.

<books>



The only piece of XML allowed outside the root element are comments and the processing instruction

<book>

<title>Cold Days</title>

<author>Jim Butcher</author>

...

</book>

</books>

XML Declaration

- Located at the very beginning – notes the version of XML you are using
- It is a processing instruction – not an element (no closing tag required)
- Before anything else
- Begin with the opening of the processing instruction(<? ?>)
- Version indicates the current version of XML to use (1.0)

```
<?xml version="1.0" ?>
```



XML Declaration

- Character encoding is the method used to represent the numbers in a character code digitally
- Can be included in your XML declaration
 - `<?xml version = "1.0" encoding = "UTF-8" ?>`
- If the encoding is NOT specified, UTF-8 or UTF-16 is assumed
- Only one encoding can be used for an entire XML document



XML Declaration

- The `standalone` attribute must be set to `yes` or `no` if used
- `Yes` specifies that the document exists entirely on its own without depending on any other files
- `No` indicates that the document *may* depend on a DTD (Document Type Definition)

```
<?xml version = "1.0" encoding = "UTF-8" standalone = "yes" ?>
```

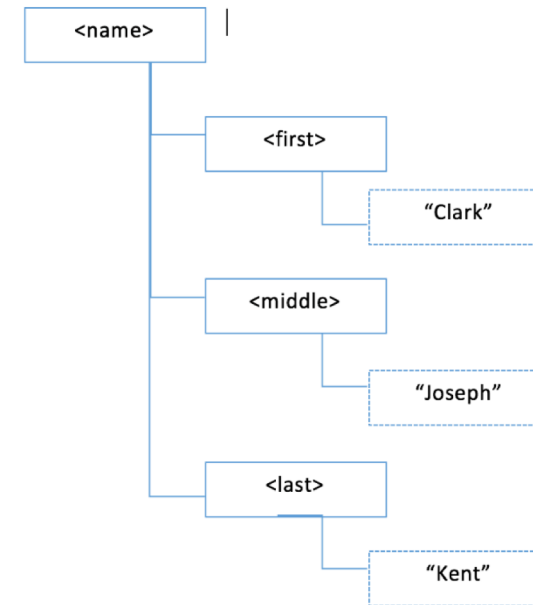
- `No` is assumed if not declared



XML hierarchy

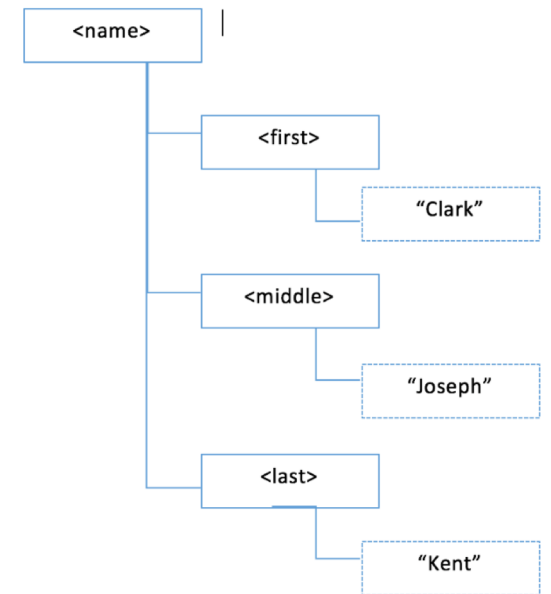
- XML groups information in hierarchies – the items relate to each other in parent/child and sibling/sibling relationships

```
<name>  
  <first>Clark</first>  
  <middle>Joseph</middle>  
  <last>Kent</last>  
</name>
```



XML hierarchy

- Also called a tree
 - Any part of the tree that contain children are called branches
 - Parts that have no children are called leaves
- Parent- child
 - `<name>` is the parent of `<first>`, `<middle>` & `<last>`
 - `<first>`, `<middle>`, `<last>` are siblings
 - The text (Clark) is a child of `<first>`



Writing XML - Elements

Must begin with a letter, or the dash (-) character, but no other numbers or punctuation – after that numbers, hyphens and decimals are allowed

No blank spaces

Can't begin with xml (either uppercase, lowercase or mixed)

Case Sensitive

Opening must match closing

Elements can contain other elements OR text (mixed content)

The name used should describe the elements purpose and its contents

