



# XML

`<an_introduction />`



# What is XML?

XML stands for eXtensible Markup Language and is a standard for coding data in a text format.

Text, because it is readable by both humans and machines.

Used for storing and transporting data.

- web services
- configuration files

# Why should I learn about XML

Why shouldn't you?

Just kidding. XML is used heavily by applications and programming technologies. Some examples:

- Web technologies like AJAX and XHTML
- Java properties files
- Build tools
- Application server configurations

On my system (Ubuntu) there are 5092 XML files...

# What does XML look like?

```
<?xml version="1.0" encoding="UTF-8"?>
<STUDENTS>
  <STUDENT id="1">
    <NAME>Anders Andersson</NAME>
  </STUDENT>
  <STUDENT id="2">
    <NAME>Beata Bengtsson</NAME>
  </STUDENT>
  <STUDENT id="3">
    <NAME>Charlie Christensen</NAME>
  </STUDENT>
  <STUDENT id="4">
    <NAME>Dick Dale</NAME>
  </STUDENT>
  <STUDENT id="5">
    <NAME>Edward Eriksson</NAME>
  </STUDENT>
</STUDENTS>
```

# Parts of an XML document

```
<?xml version="1.0" encoding="UTF-8"?>
<STUDENTS>
  <STUDENT id="1">
    <NAME>Anders Andersson</NAME>
  </STUDENT>
  <STUDENT id="2">
    <NAME>Beata Bengtsson</NAME>
  </STUDENT>
  <STUDENT id="3">
    <NAME>Charlie Christensen</NAME>
  </STUDENT>
  <STUDENT id="4">
    <NAME>Dick Dale</NAME>
  </STUDENT>
  <STUDENT id="5">
    <NAME>Edward Eriksson</NAME>
  </STUDENT>
</STUDENTS>
```

- <- “Prolog” (XML version and character encoding)
- <- Root element
- <- Child element
- <- Text content (all elements can have it!)
  
- <- id="2" is an “Attribute”
  
- <- tags are closed like this
  
- <- <NAME> is a child to <STUDENT id="3">

# Structural rules

- All XML documents must have a root element
- All elements have an opening tag and a closing tag
- `<tag>some text contents if any</tag>`
- All elements can have child elements
- All elements must be “closed”
- UTF-8 is the default encoding used
- Element names are case sensitive
- Attribute values must be quoted (single or double quotes work)

# Rules for the text content

For the text contents of an element, use these entities:

<b>Use</b>	<b>instead of</b>	<b>meaning</b>
&lt;	<	less than
&gt;	>	greater than
&amp;	&	ampersand
&apos;	'	apostrophe
&quot;	"	quotation mark

You can use `<!-- Comments like this -->` for contents not holding data

# Elements are not predefined

There is no list of elements for XML (as for HTML for instance).

XML is extensible and general purpose. You make up the element names.

White space counts! `<name>H e n r i k</name>` means “H e n r i k”.

Lines end with LF (and nothing else)



# Well-formed documents

Documents following the syntax rules are said to be well-formed.

Well-formedness is only about structure and syntax. Never about data or meaning.

Many applications refuse to even try to interpret (parse) documents which are not well-formed.

This allows for some level of data integrity.

# Empty elements

These two are equivalent:

`<name></name>`      `<name />`

# Metadata is often coded in attributes

Metadata is data about the data (and not about the real world).

Often, attributes are used for metadata, but there are no rules for this.

```
<student id="1">  
  <name>Lena Andersson</name>  
</student>
```

Above, id is a piece of metadata (students don't have IDs in the real world).  
Not in Sweden, anyway.

# XML is not HTML

HTML is a markup language created for displaying data.

XML doesn't contain information about how to display the document. It is solely a carrier for data.

# But XHTML is XML

XHTML is a standard for web content. It is a subset of XML.

This is just one of millions of examples of subsets of XML.

Remember, XML is just a carrier for data, so it is all about syntax and structure.

The extensible nature of XML has made it popular for many applications.

# Further reading

<http://www.w3schools.com/xml/default.asp>

<https://www.tutorialspoint.com/xml/>

<http://www.javaworld.com/article/2076396/java-xml/xml-for-the-absolute-beginner.html>

<http://www.xmlmaster.org/en/article/d01/>