

1 XML

XML (eXtensible Markup Language) is a language with many uses. One of them is to transport data between different systems.

XML consists of two languages, one language for the actual XML documents and one language for specifying how the XML documents¹ should be structured, called DTD (Document Type Definition). DTD is the older language for defining XML structures. Another “newer” language is XML Schema, which is somewhat more powerful than DTD. Not all XML documents are associated to DTDs or XML Schemas. Here is an example of an XML document and its DTD:

XML Document (saved in a file called “book.xml”)

```
<?xml version="1.0"?>
<!DOCTYPE book SYSTEM "d:\dtd\book.dtd">
<book>
  <chapter id="1" date="07/01/1997">
    <section>This is a section in Chapter One.</section>
  </chapter>
  <chapter id="2" date="01/02/1997">
    <section>This is a section in Chapter Two.</section>
    <footnote>A footnote in Chapter Two is here.</footnote>
  </chapter>
  <price date="1998-12-22" time="11:12:13">38.281</price>
</book>
```

DTD (file “book.dtd”)

```
<!ELEMENT book (author*,chapter*,price)>
<!ELEMENT author (#PCDATA)>
<!ELEMENT chapter (section*, footnote*)>
<!ATTLIST chapter id (1|2|3) #REQUIRED
               date CDATA #IMPLIED>
<!ELEMENT price (#PCDATA)>
<!ATTLIST price date CDATA #IMPLIED
               time CDATA #IMPLIED>
<!ELEMENT section (#PCDATA)>
<!ELEMENT footnote (#PCDATA)>
```

❗ Both languages are case sensitive!

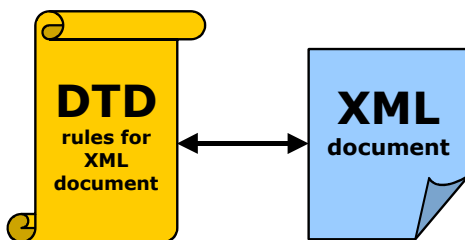


Figure 1 XML and DTD

An XML document can refer to a DTD file. A DTD file can be associated with many XML documents. When an XML document refers to a DTD file then the XML document's content is supposed to follow the rules defined in the DTD file.

1.1 XML documents

Elements:

In the previous example **chapter** is an element. Everything from the **<chapter>** to the **</chapter>** constitutes an element **chapter**.

Every XML document must have a root element, an element that has its start tag in the beginning of the XML document and its end tag at the end of the XML document. This element may appear only once in the XML document.

¹ The term XML document refers to a file often with the extension .xml.

Attributes:

The element **chapter** has an attribute **id** and an attribute **date**. All attributes of an element appear within the start tag of the element. Attributes have a value that is within double quotation marks (“”).

Structure:

```
<element attribute1="value" attribute2="value2">  
    element content  
</element>
```

The element content can be empty, text or other elements.
If the element content is empty then the element can look like this:

```
<element attribute1="value" attribute2="value2" />
```

If an end tag is used then no characters are allowed between the start tag and the end tag:

```
<element attribute1="value" attribute2="value2"></element>
```

XML declaration & DOCTYPE element

The first two lines of any XML document are normally the XML declaration & the DOCTYPE declaration:

XML declaration:

```
<?xml version="1.0" standalone="no"?>
```

In the XML declaration we define the XML version and whether there is a DTD or XML Schema file with rules for the XML structure or not.

DOCTYPE declaration:

```
<!DOCTYPE Book SYSTEM "d:\dtd\book.dtd">
```

The DOCTYPE defines the root element of the XML document and the SYSTEM points out the DTD or XML Schema file for the XML document.

1.2 DTD

The DTD file contains rules to be followed when constructing an XML document.

It defines the elements that can appear in the XML document:

```
<!ELEMENT element-name>
```

It defines the elements that can appear within an element:

```
<!ELEMENT element-name (element2-name)>
```

or the type of the element content:

```
<!ELEMENT element-name (#PCDATA)>
```

It also defines the attributes that an element can have, with the appropriate rules (the type of the attribute, whether it has to be there or not, a default value, etc.):

```
<!ATTLIST element-name
  attribute1-name CDATA #REQUIRED
  attribute2-name CDATA #IMPLIED>
```

1.3 XML Schema

XML Schema is a new and improved approach to defining rules for XML documents. Unlike DTD, XML Schema uses XML syntax. With XML Schema, one can define more complex details about the structure of XML documents and the allowed values.

XML Schemas have the following root element:

```
<schema xmlns="http://www.w3.org/2001/XMLSchema">
```

Elements are defined with the element "element". The structure of the defined element is specified by a referred type:

```
<element name="Book" type="BookType"/>
```

The types can be defined using either "complexType" or "simpleType", if not a simple data type. A complexType can be suitable to define subelements and attributes:

```
<complexType name="BookType">
  <sequence>
    <element name="Author" type="AuthorType" minOccurs="0" maxOccurs="unbounded"/>
    <element name="Chapter" type="ChapterType" minOccurs="0" maxOccurs="unbounded"/>
    <element name="Price" type="PriceType"/>
  </sequence>
</complexType>

<complexType name="PriceType">
  <attribute name="time" type="time" use="optional"/>
  <attribute name="date" type="date" use="optional"/>
</complexType>
```

1.4 Links

For more help on how to construct an XML document visit one of the following tutorial sites (tutorials for XML, DTD and XML Schema):

- <https://www.w3schools.com/xml/default.asp>
- https://www.w3schools.com/xml/xml_dtd_intro.asp
- https://www.w3schools.com/xml/schema_intro.asp