Return to Module Overview Page (https://onlinelearning.berkeley.edu/courses/665040/pages/module-nine)

9.3: The Document Type Definition (DTD)

As important as XML files are, DTD files are equally as essential.

DTD stands for Document Type Definition. A DTD specifies in an XML document:

- element names
- element type
- element attributes
- attribute type
- element nesting
- order of element
- number of elements (0 or 1, 0 or more, 1 or more)

An XML document that follows the specification of a DTD is called a valid document. A document must be well-formed and valid if it is to parse correctly.

In the example below, we define a DTD (dvd.dtd) that will validate our example XML document.

```
<!ELEMENT DVDS (DVD*) >

<!ELEMENT DVD (title, performers, discs, price) >

<!ATTLIST DVD id CDATA #REQUIRED>

<!ELEMENT title (#PCDATA)>

<!ELEMENT performers (performer*)>

<!ELEMENT performer (#PCDATA)>

<!ELEMENT discs (#PCDATA)>

<!ELEMENT price (#PCDATA)>
```

The first line defines the root element called DVDS. It nests the DVD element as indicated by the surrounding parenthesis (). The * symbol specifies that it can have zero or more DVD elements.

The second line specifies that the DVD element can accept the following elements: title, performers, discs and price. (lines 4, 5, 7 and 8)

The third line specifies that DVD declares an attribute named id and that it is a CDATA (or character data) type. This attribute is indicated to be required by putting the #REQUIRED keyword at the end.

The fifth line specifies that performers can have zero or more performer elements. The performer element must be a #PCDATA (or parsed character data) type.

The remaining lines (title, discs, price) define the other sub elements of DVD. #PCDATA is similar to CDATA and means that the character data is parsed. Use for parsing character data within tags.