

XML Technologies

XML, DTD

Basic Structural Components of XML Documents (1/2)

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- XML declaration -->
<!-- root element -->
<root-element>

    <!-- an empty element -->
    <element/>

    <!-- attributes of an element -->
    <element attribute1="value"
              attribute2="another value">...</element>

    <!-- an element with subelements -->
    <element attribute1="value" attribute2="value">
        <subelement>TEXT CONTENT</subelement>
        <subelement>...</subelement>
    </element>
</root-element>
```

Basic Structural Components of XML Documents (2/2)

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

```
<!-- XML declaration -->
```

```
<!-- root element -->
```

```
<root-element>
```

```
  <!-- element with mixed content -->
```

```
  <element>
```

```
    TEX <subelement>MIXED</subelement> CONTENT
```

```
    WITH <subelement>SUBELEMENTS</subelement>
```

```
  </element>
```

```
</root-element>
```

Common Mistakes in XML Documents

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

```
<elementA>
```

```
  <!-- a bad enclosing symbol -->
```

```
  <elementB>...</elementb>
```

```
  <!-- an element without an enclosing symbol -->
```

```
  <elementD>
```

```
  <!-- a bad element nesting -->
```

```
  <elementE>
```

```
    <elementF>
```

```
      ...
```

```
    </elementE>
```

```
  </elementF>
```

```
</elementA>
```

```
<!-- another root element -->
```

```
<elementG />
```

Basic Structural Components in DTD (1/3)

<!-- declaration of an element with text content -->

<!ELEMENT AAA (#PCDATA)>

<!-- declaration of an element with empty content -->

<!ELEMENT AAA EMPTY>

<!-- declaration of an element with any content -->

<!ELEMENT AAA ANY>

<!-- declaration of an element with a sequence of subelements -->

<!ELEMENT AAA (BBB,CCC?,DDD*,EEE+)>

<!-- declaration of an element with a choice of subelements -->

<!ELEMENT AAA (BBB|CCC)>

<!-- declaration of an element with a complex content model -->

<!ELEMENT AAA (BBB, ((CCC,DDD*)|EEE+))>

<!-- declaration of an element with a mixed content -->

<!ELEMENT AAA (#PCDATA|BBB|CCC|DDD)*>

Basic Structural Components in DTD

(2/3)

```
<!-- declaration of attributes of element AAA :  
  - a mandatory attribute-1 with text content  
  - an optional attribute-2 with text content  
  - attribute-3 with a fixed "value"  
  - a key attribute-4  
  - attribute-5 is a key reference  
  - attribute-6 has one of the values, "value1" is default (if the  
attribute is not present in XML document, it has this value) -->
```

```
<!ATTLIST AAA
```

```
attribute-name-1 CDATA #REQUIRED
```

```
attribute-name-2 CDATA #IMPLIED
```

```
attribute-name-3 CDATA #FIXED "value"
```

```
attribute-name-4 ID ...
```

```
attribute-name-5 IDREF ...
```

```
attribute-name-6 (value1|value2|...) "value1">
```

Basic Structural Components in DTD

(3/3)

```
<!-- internal entity declaration -->
```

```
<!ENTITY lt "&#38;#60;">
```

```
<!ENTITY c "Copyright: " >
```

```
<!-- declaration of system (private, own) external entity -->
```

```
<!ENTITY ct SYSTEM "http://ksi.mff.cuni.cz/copyright.xml" >
```

```
<!-- declaration of public external entity -->
```

```
<!ENTITY ct PUBLIC "-//W3C//TEXT copyright//EN"
```

```
"http://www.w3.org/xmlspec/copyright.xml">
```

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

```
<copyright>&c;&ct;</copyright>
```

DTD Declaration

```
<!-- internal declaration -->
```

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

```
<!DOCTYPE root-element [...DTD...]>
```

```
<root-element>...</root-element>
```

```
<!-- private, external DTD -->
```

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

```
<!DOCTYPE root-element SYSTEM "URI">
```

```
<root-element>...</root-element>
```

```
<!-- public, external DTD -->
```

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

```
<!DOCTYPE root-element PUBLIC public-name>
```

```
<root-element>...</root-element>
```

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

```
<!DOCTYPE root-element PUBLIC public-name "URI">
```

```
<root-element>...</root-element>
```


Keys and References

```
<!ATTLIST product vid ID #REQUIRED>
<!ATTLIST order oid ID #REQUIRED>
<!ATTLIST item vid_ref IDREF #REQUIRED>

<sale>
  <order oid="id_1">
    <item vid_ref="id_1" />
    <item vid_ref="id_3" />
  </order>
  <order oid="id_2">
    <item vid_ref="id_2" />
    <item vid_ref="id_4" />
    <!-- the next reference leads to non existing id! -->
    <item vid_ref="id_5" />
  </order>
  <product vid="id_1" /><!-- ! "id_1" is already used -->
  <product vid="id_2" /><!-- ! "id_2" is already used -->
  <product vid="id_3" />
  <product vid="id_4" />
</sale>
```

Common Mistakes in DTD (1/2)

```
<!-- Every element which we use in a DTD must be defined. The same holds for attributes. -->
```

```
<!-- usage of #PCDATA for different purpose than text or mixed content -->
```

```
<!ELEMENT AAA (BBB, #PCDATA, CCC)>
```

```
<!-- the value of ID attribute starts with number -->
```

```
<!ATTLIST BBB number ID #REQUIRED>
```

```
<BBB number="1">TEXT</BBB>
```

```
<!-- EMPTY can be used only separately -->
```

```
<!ELEMENT AAA (EMPTY | (BBB,CCC) )>
```

```
<!-- attributes can not be used in element declaration -->
```

```
<!ELEMENT AAA (attr1 | (attr2,attr3))>
```

```
<!ATTLIST AAA attr1 CDATA  
                attr2 CDATA  
                attr3 CDATA>
```

Common Mistakes in DTD (2/2)

<!-- Only elements declared in the DTD can be used in the XML document (in the following example the element customer is not defined) -->

```
<!ELEMENT new-customer (...)>
```

```
<!ELEMENT present-customer (...)>
```

```
<customer>...</customer>
```

<!-- An attribute declared as choice must have a default value or must be compulsory -->

```
<!ATTLIST AAA attr1 (value1|value2|value3)>
```

<!-- It is not possible to specify the content of the root element directly in DOCTYPE. It must be done in declaration of this element within the DTD. -->

```
<!DOCTYPE AAA (BBB,CCC) [...DTD...]>
```