Web Technology
(KCS-602)
Unit 2
DTD

Prepared By

Abhishek Kesharwani

Assistant Professor,UCER Naini,Allahabad

Introduction to DTD

- A Document Type Definition (DTD) defines the legal building blocks of an XML document. It defines the document structure with a list of legal elements and attributes.
- A DTD can be declared inline inside an XML document, or as an external reference.

Why Use a DTD?

- With a DTD, each of your XML files can carry a description of its own format.
- With a DTD, independent groups of people can agree to use a standard DTD for interchanging data.
- Your application can use a standard DTD to verify that the data you receive from the outside world is valid.
- You can also use a DTD to verify your own data.

DTD - XML Building Blocks

PCDATA

- PCDATA means parsed character data.
- Think of character data as the text found between the start tag and the end tag of an XML element.
- **PCDATA is text that WILL be parsed by a parser**. **The text will be examined by the parser for entities and markup**.
- Tags inside the text will be treated as markup and entities will be expanded.
- However, parsed character data should not contain any &, <, or > characters; these need to be represented by the & < and > entities, respectively.

CDATA

- CDATA means character data.
- **CDATA is text that will NOT be parsed by a parser**. Tags inside the text will NOT be treated as markup and entities will not be expanded.

Declaring Elements

- In a DTD, XML elements are declared with an element declaration with the following syntax:
- <!ELEMENT element-name category>
  or
  <!ELEMENT element-name (element-content)>

Empty Elements

- Empty elements are declared with the category keyword EMPTY:
- <!ELEMENT element-name EMPTY>

Example:

<!ELEMENT br EMPTY>
XML example:

Elements with Parsed Character Data

- Elements with only parsed character data are declared with #PCDATA inside parentheses:
- <!ELEMENT element-name (#PCDATA)>

Example:

<!ELEMENT from (#PCDATA)>

Elements with any Contents

- Elements declared with the category keyword ANY, can contain any combination of parsable data:
- <!ELEMENT element-name ANY>

  Example:

  <!ELEMENT note ANY>

Elements with Children (sequences)

- Elements with one or more children are declared with the name of the children elements inside parentheses:
- <!ELEMENT element-name (child1)>
  or
  <!ELEMENT element-name (child1,child2,...)>

  Example:

  <!ELEMENT note (to,from,heading,body)>

Declaring Only One Occurrence of an Element

- <!ELEMENT element-name (child-name)>

  Example:

  <!ELEMENT note (message)>

Declaring Minimum One Occurrence of an Element

<!ELEMENT element-name (child-name+)>

Example:

<!ELEMENT note (message+)>

Declaring Zero or More Occurrences of an Element

<!ELEMENT element-name (child-name*)>

Example:

<!ELEMENT note (message*)>

Declaring either/or Content

Example:

<!ELEMENT note (to,from,header,(message|body))>

The example above declares that the "note" element must contain a "to" element, a "from" element, a "header" element, and either a "message" or a "body" element

TYPES of DTD

- Internal DTD
- External DTD

Internal DTD Declaration

If the DTD is declared inside the XML file, it should be wrapped in a DOCTYPE definition with the following syntax:

<!DOCTYPE root-element [element-declarations]>

Example XML document with an internal DTD:

```
<?xml version="1.0"?>
<!DOCTYPE note [
<!ELEMENT note (to,from,heading,body)>
<!ELEMENT to (#PCDATA)>
<!ELEMENT from (#PCDATA)>
<!ELEMENT heading (#PCDATA)>
<!ELEMENT body (#PCDATA)>
]>
```

```
<note>
<to>Tove</to>
<from>Jani</from>
<heading>Reminder</heading>
<body>Don't forget me this
weekend</body>
</note>
```

- **!DOCTYPE note** defines that the root element of this document is note
- **!ELEMENT note** defines that the note element contains four elements: "to,from,heading,body"
- **!ELEMENT to** defines the to element to be of type "#PCDATA"
- **!ELEMENT from** defines the from element to be of type "#PCDATA"
- **!ELEMENT heading** defines the heading element to be of type "#PCDATA"
- **!ELEMENT body** defines the body element to be of type "#PCDATA"

External DTD Declaration

- If the DTD is declared in an external file, it should be wrapped in a DOCTYPE definition with the following syntax:
- <!DOCTYPE root-element SYSTEM "filename">
- <?xml version="1.0"?>
  <!DOCTYPE note SYSTEM "note.dtd">
  <note>
  <to>Tove</to>
  <from>Jani</from>
  <heading>Reminder</heading>
  <body>Don't forget me this weekend!</body>
  </note>

And this is the file "note.dtd" which contains the DTD:

```
<!ELEMENT note (to,from,heading,body)>
<!ELEMENT to (#PCDATA)>
<!ELEMENT from (#PCDATA)>
<!ELEMENT heading (#PCDATA)>
<!ELEMENT body (#PCDATA)>
```