

# Il linguaggio XML nell'ambito della catalogazione di immagini

DI ANNA CASCONI E RITA COCCO



Si modelli un documento XML di catalogazione Libri e relativo DTD di validazione in cui:

- ▶ Un Catalogo può contenere zero o più Libri ;
- ▶ Un Libro è descritto da un Titolo, almeno un Autore, zero o più Protagonisti ed eventualmente un Genere ;
- ▶ Un **Libro** è dotato di proprietà quali un codice identificativo univoco ( **SBN** ) , un'indicazione del tipo di **formato** ( obbligatoria ) che può essere **cartaceo**, **ebook**, **audiolibro**, un'indicazione dello **stato** ( **nuovo** | **sottolineato** | **danneggiato** ) , un'indicazione sulla prestabilità ( **prestabile** - sì o no ).

Ecco un documento XML valido secondo il DTD appena definito:

```
<?xml version = " 1.0 " > encoding = "utf-8" ?>
< !DOCTYPE Catalogo SYSTEM "Catalogo.dtd" >
<Catalogo>
  <Libro SBN = "978-88-06-20282-8" formato = cartaceo >
    <Titolo> Vizio di Forma </Titolo>
    <Autore> Thomas Pynchon </Autore>
    <Protagonista> Doc Sportello </Protagonista>
    <Genere> Noir </Genere>
  </Libro>
  <Libro SBN = "978-88-07-88279-1" formato = cartaceo >
    <Titolo> Il lungo addio </Titolo>
    <Autore> Raymond Chandler </Autore>
    <Protagonista> Philip Marlowe </Protagonista>
    <Genere> Noir </Genere>
  </Libro>
  ...
</Catalogo>
```

## Questo è il DTD completo:

```
<! ELEMENT Catalogo ( Libro* ) >
<! ELEMENT Titolo ( #PCDATA ) >
<! ELEMENT Autore ( #PCDATA ) >
<! ELEMENT Protagonista ( #PCDATA ) >
<! ELEMENT Genere ( #PCDATA ) >
<! ELEMENT Libro ( Titolo, Autore+ , Protagonista*, Genere? ) >
<! ATTLIST Libro SBN #REQUIRED
  prestabile ( sì | no ) ' sì '
  formato ( cartaceo | ebook | audiolibro ) ' cartaceo '
  stato ( nuovo | sottolineato | danneggiato ) ' nuovo ' >
```

# DTD: Document Type Definition

- ▶ E' un linguaggio per definire la grammatica che descrive la composizione degli elementi costituenti una certa classe di documenti XML ;
- ▶ Fornisce uno strumento per la validazione dei documenti XML ;
- ▶ E' costituito da un elenco di dichiarazioni che descrivono la struttura del documento e ne definiscono gli elementi strutturali ( **element** ) , il modello di contenuto di ogni elemento ( **content model** ) e gli **attributi** ad essi associati.

Per applicare un DTD ad un documento XML nel suo prologo dobbiamo inserire una dichiarazione – che va posta sotto l'XML Declaration – con questa sintassi:

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

dove:

- root-element è il nome dell'elemento radice ;
- system è il documento di utilizzo locale ( se pubblico va sostituito da Public )
- "filename" è il file che contiene il DTD.

Per dichiarare un elemento si usa la sintassi:

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

E per gli attributi invece:

```
<!ATTLIST ElementName  
    AttrName1 AttrType1 Value1  
    AttrName2 AttrType2 Value2  
... >
```

# Com'è fatto un documento XML?

- ▶ E' strutturato in modo gerarchico e composto da elementi: un **elemento radice** e dei **sotto-elementi** eventualmente contenuti nel primo ;
- ▶ Ad un elemento possono essere associate informazioni descrittive chiamate **attributi** ;
- ▶ Ogni documento XML può essere rappresentato come un albero ( **document tree** ) .

# Struttura formale di un documento XML

Un documento XML è costituito da due parti

- un **prologo**, che contiene la dichiarazione XML ed il riferimento ( opzionale ) ad altri documenti che ne definiscono la struttura o direttive di elaborazione ;
- il **corpo**, che è il documento XML vero e proprio.



# Elementi e tag

- ▶ Un elemento è un frammento di testo racchiuso fra uno **start tag** e un **end tag**:

- ▶ Uno **start tag** è costituito da un nome + eventuali attributi racchiusi dai simboli '<' '>'

<TagName attribute-list > ES. <Libro SBN = "978-88-06-20282-8" formato = cartaceo >

- ▶ Un **end tag** è costituito dal nome dello stesso start tag racchiuso da '</' '>'

</TagName> ES. </Libro>