

Definizione del tipo di documento (DTD)

Anche se non necessario, è sempre molto utile, in particolare quando si utilizzano strutture complesse, definire in maniera precisa la struttura dei dati prima di costruire il file XML.

A questo scopo si usa una **DTD** (*Document Type Definition*) che consente di definire in modo non ambiguo la struttura dei dati trattati.

Il codice DTD può essere inserito direttamente all'interno del file XML, oppure può essere salvato in un file con estensione **.dtd** che viene poi collegato al file XML.

Progetto

Definire il tipo di documento per l'organizzazione dei *bookmark* di siti Web.

La struttura gerarchica che si vuole ottenere è la seguente: *lista* comprende uno o più dati di tipo *sito* che, a loro volta, comprendono uno o più dati semplici *nome*, *URL*, *descrizione* e *categoria*.



La definizione del tipo di documento per l'esempio precedente è la seguente:

```
<!DOCTYPE lista[
<!ELEMENT lista (sito+)>
<!ELEMENT sito (nome, URL, descrizione, categoria)>
<!ELEMENT nome (#PCDATA)>
<!ELEMENT URL (#PCDATA)>
<!ELEMENT descrizione (#PCDATA)>
<!ELEMENT categoria (#PCDATA)>
<!ATTLIST sito IDSito ID #REQUIRED>
]>
```

La definizione del documento deve essere scritta usando la sintassi indicata con i segni <, >, ! e parentesi quadre di apertura e chiusura.

È importante sottolineare che questo codice non contiene alcun dato, ma determina solamente la struttura dei dati che vengono rappresentati.

La clausola **!DOCTYPE** definisce l'oggetto *lista*, formato da vari elementi indicati con **!ELEMENT**. Il primo elemento *lista* è la radice della struttura che contiene elementi di tipo *sito*. Nella parentesi tonda, dopo il nome *sito*, è stato scritto il segno + per indicare che l'elemento *lista* può contenere uno o più elementi *sito*.

L'elemento *sito* è formato da quattro dati indicati tra parentesi tonde.

Ogni elemento può essere formato da un singolo dato o da un dato strutturato, permettendo così la creazione di strutture complesse. In generale è possibile definire un dato come facoltativo o obbligatorio, di tipo testo o numerico.

Il tipo **#PCDATA** (*Parsed Character Data*) indica un dato di tipo testo (cioè formato da caratteri) che può essere analizzato da un parser.

Il **parser**, in generale, è un elaboratore di testi che possiedono una struttura sintattica predefinita. I parser XML elaborano un file XML per estrarne i dati passandoli a un browser Web oppure a un'applicazione. I moderni browser, come *Internet Explorer*, *Firefox* o *Chrome*, sono dotati di un parser che consente di visualizzare un file XML.

La clausola **!ATTLIST** definisce un attributo per l'elemento sito: l'attributo è *IDSito*, di tipo **ID**, cioè il suo valore deve essere univoco all'interno del documento ed è un attributo obbligatorio (**#REQUIRED**).

I valori degli attributi di tipo ID non possono iniziare con un numero.

La definizione DOCTYPE del tipo di documento può essere inserita direttamente nel file XML oppure può essere salvata in un file separato con l'estensione **.dtd**. In questo secondo caso, il codice che rappresenta la struttura del documento è il seguente (non si scrive la parte *!DOCTYPE*):

(siti.dtd)

```
<!ELEMENT lista (sito+)>
<!ELEMENT sito (nome, URL, descrizione, categoria)>
<!ELEMENT nome (#PCDATA)>
<!ELEMENT URL (#PCDATA)>
<!ELEMENT descrizione (#PCDATA)>
<!ELEMENT categoria (#PCDATA)>
<!ATTLIST sito IDSito ID #REQUIRED>
```

A questo punto è necessario collegare il file *siti.dtd* al file principale inserendo, nel file XML, la riga

```
<!DOCTYPE lista SYSTEM "siti.dtd">
```

Dopo la clausola *!DOCTYPE* occorre indicare il nome dell'elemento che rappresenta il livello più alto (la radice) della struttura dei dati.

Il file XML, con l'aggiunta del riferimento al documento *.dtd*, è il seguente:

(siti1.xml)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE lista SYSTEM "siti.dtd">
<lista>
  <sito IDSito="s001">
    <nome>Google</nome>
    <URL>http://www.google.it</URL>
    <descrizione>Il piu' famoso motore di ricerca</descrizione>
    <categoria>motori di ricerca</categoria>
  </sito>
  <sito IDSito="s002">
    <nome>La Repubblica</nome>
    <URL>http://www.repubblica.it</URL>
    <descrizione>Versione on-line del quotidiano</descrizione>
    <categoria>news</categoria>
  </sito>
</lista>
```