

XML Information Set

1 XML Information Set

XML Information Set (abbreviato **XML Infoset**) è, per il W3C, il modello di riferimento per i dati semi-strutturati serializzati in XML.

Ogni documento XML ben formato (dal punto di vista della serializzazione) può essere rappresentato da un **infoset**, che è un insieme di **information item** (ciascuno dei quali corrisponde, approssimativamente, a un nodo dell'albero). XML Infoset definisce ben 11 diversi tipi di information item, e a ciascuno di essi sono associate diverse proprietà; i principali sono:

Document Information Item: costituisce la radice dell'albero. Esso *non* è un elemento del documento: è il padre dell'elemento radice, ed è presente nel modello dei dati, ma non nella serializzazione. Alcune delle sue proprietà sono:

- la versione di XML usata per la serializzazione;
- la codifica dei caratteri;
- il nodo figlio (che è l'elemento radice).

Element Information Item: corrisponde a un elemento (cioè, nella serializzazione, a una coppia di tag – uno di apertura e uno di chiusura). Le sue principali proprietà sono:

- il nome dell'elemento;
- il namespace di appartenenza;
- gli attributi;
- la lista degli information item figli;
- l'information item padre.

Attribute Information Item: corrisponde a un attributo di un elemento. Le sue proprietà più importanti sono il nome, il valore e l'elemento possessore di tale attributo. Tecnicamente, la relazione tra un elemento e un suo attributo non viene considerata una relazione padre-figlio. Infatti, i figli di un elemento sono ordinati, mentre gli attributi non si inseriscono in tale ordine (e, in generale, l'ordine degli attributi non è proprio considerato). In pratica, essi sono trattati come “entità a parte”.

Character Information Item: corrisponde a un singolo carattere di una stringa di dati (mentre, in XPath Data Model, l'intera stringa verrebbe rappresentata da un singolo nodo di tipo text).

Processing Information Item: analogo a un nodo di tipo processing di XPath Data Model.

Come per XPath Data Model, un'istanza di XML Infoset comprende sia informazioni relative alla struttura (ad albero) di un documento XML (cioè allo "schema" del documento) che informazioni relative ai dati contenuti (cioè l'"istanza" dello schema):

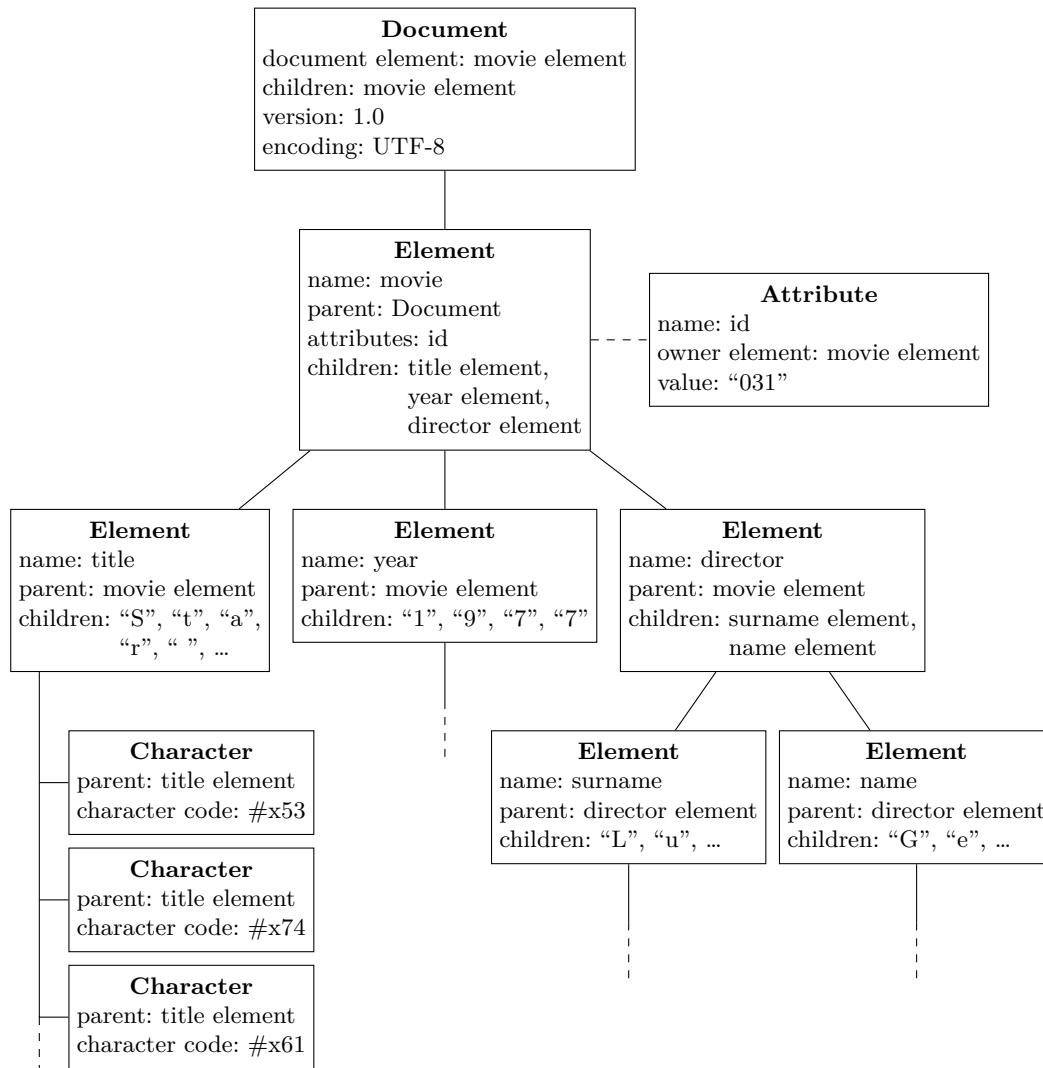
- la struttura è specificata dai nomi degli Element Information Item e dalle loro relazioni di padre/figlio (rappresentate tramite le proprietà di tali item);
- i dati sono i valori presenti nelle proprietà dei Character Information Item.

2 Esempio

Dato il seguente documento XML,

```
<?xml version="1.0" encoding="UTF-8"?>
<movie id="031">
  <title>Star Wars</title>
  <year>1977</year>
  <director>
    <surname>Lucas</surname>
    <name>George</name>
  </director>
</movie>
```

l'infoset corrispondente è:



3 Distinzione tra documento e istanza di un modello

L'infoset (o l'istanza di XPath Data Model) *non* è il documento XML. Infatti, l'informazione contenuta nell'infoset può essere materializzata come serializzazione (documento) XML, ma anche in forma tabellare, o in altre forme ancora.

4 Differenze tra XPath Data Model e XML Infoset

- XPath Data Model è più semplice rispetto a XML Infoset (ad esempio, il primo definisce 7 tipi di nodi, mentre il secondo prevede 11 tipi di information item).

- La struttura ad albero è più esplicita in XPath Data Model, mentre XML Infoset la rappresenta mediante le proprietà padre/figli degli information item.
- La rappresentazione dei dati “veri e propri” (le foglie dell’albero) è più semplice in XPath Data Model, che prevede un singolo text node per ogni stringa: in XML Infoset, una stringa diventa invece una sequenza di Character Information Item consecutivi.