

# Integrazione di XML in SQL

## 1 XMLTABLE

XMLTABLE è una delle funzioni predefinite di SQL/XML che permettono di estrarre dati relazionali da alberi XML (tipicamente, ma non necessariamente, contenuti in delle tabelle) – il “contrario” delle funzioni di publishing. Per illustrarne il funzionamento, si presentano in seguito due esempi, riferiti alla tabella `movies_xml`:

```

id  movie_xml
-----
001 <movie>
    <title>...</title>
    <running-time>...</running-time>
    <year-released>...</year-released>
    <director>
        <family-name>...</family-name>
        ...
    </director>
    ...
</movie>
002 <movie>
    <title>...</title>
    <running-time>...</running-time>
    <year-released>...</year-released>
    <director>
        <family-name>...</family-name>
        ...
    </director>
    <director>
        <family-name>...</family-name>
        ...
    </director>
    ...
</movie>

```

- Estrarre titolo, lunghezza e anno di ciascun film, ottenendo come risultato una tabella “piatta”, senza XML.

```

SELECT result.*
FROM
  movies_xml,
  XMLTABLE(
    'for $m in $col/movie
    return $m'
    PASSING movies_xml.movie_xml AS "col"
    COLUMNS
      "title" VARCHAR(80) PATH 'title',
      "running_time" INTEGER PATH 'running-time',
      "year_released" INTEGER PATH 'year-released'
  ) AS result;

```

Gli argomenti di XMLTABLE, che specificano come costruire una relazione a partire da un albero XML, si possono considerare suddivisi in due parti:

– Il *row pattern*,

```

'for $m in $col/movie
return $m'
PASSING movies_xml.movie_xml AS "col"

```

indica l'espressione XQuery che viene valutata per generare ciascuna riga della relazione risultante.

– Il *column pattern*,

```

COLUMNS
  "title" VARCHAR(80) PATH 'title',
  "running_time" INTEGER PATH 'running-time',
  "year_released" INTEGER PATH 'year-released'

```

determina nomi, tipi e valori degli attributi della relazione risultante. In particolare, il valore di ogni attributo è specificato scrivendo, dopo la parola chiave PATH, un'espressione XPath (o XQuery) che viene valutata nel contesto del risultato del row pattern

Il risultato di quest'interrogazione sarà una tabella piatta come la seguente:

title	running_time	year_released
Star Wars	120	1977
...	...	...

- Calcolare la lunghezza media dei film di ciascun regista<sup>1</sup> usciti in ciascuno degli anni 2000, 2001 e 2002, e la lunghezza media complessiva per ciascun anno (su tutti i registi).

---

<sup>1</sup>Supponendo che un film possa avere più registi, qui si considera solo quello che compare per primo nel documento XML.

```

SELECT
  result.year_released,
  result.director,
  AVG(result.running_time) AS average_length
FROM
  movies_xml,
  XMLTABLE(
    'for $m in $col/movie
    return $m'
    PASSING movies_xml.movie_xml AS "col"
    COLUMNS
      "year_released" INTEGER PATH 'year-released',
      "running_time" INTEGER PATH 'running-time',
      "director" VARCHAR(80) PATH 'director[1]/family-name'
  ) AS result
WHERE result.year_released IN (2000, 2001, 2002)
GROUP BY ROLLUP(result.year_released, result.director);

```

Qui si usano, in combinazione, le funzionalità XML e OLAP di SQL, che consentono di effettuare analisi anche relativamente avanzate su dati semi-strutturati, direttamente all'interno del DBMS.

Il risultato della query ha la seguente forma:

year_released	director	average_length
2000	...	...
2000	...	...
2000	...	...
2000	NULL	...
2001	...	...
2001	...	...
2001	NULL	...
2002	...	...
2002	...	...
2002	...	...
2002	...	...
2002	NULL	...
NULL	NULL	...

## 2 Aspetti non considerati dallo standard

Uno degli aspetti dell'interazione tra il modello relazionale e i modelli semi-strutturati che non viene preso in considerazione dallo standard SQL/XML è la definizione di una

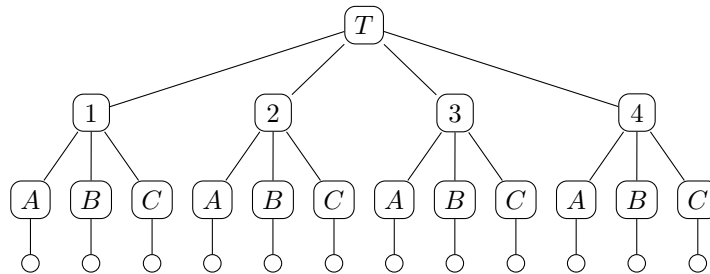
metodologia di progettazione che “guidi” il mapping tra una rappresentazione ad albero e una rappresentazione tabellare dei dati, e viceversa.

Ad esempio, una tabella  $T$  con gli attributi  $A, B, C$  e le righe (tuple) 1, 2, 3, 4

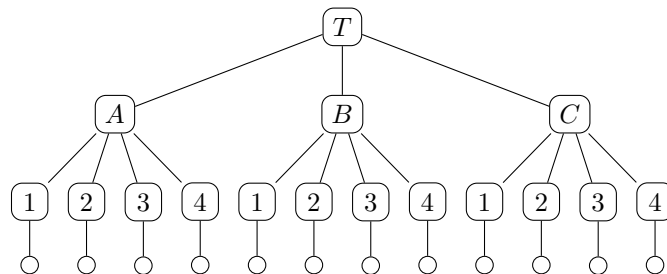
$A$	$B$	$C$	
			1
			2
			3
			4

può corrispondere ad almeno due alberi diversi:

- uno che rappresenta prima le righe, e poi le colonne:



- uno che, al contrario, rappresenta prima le colonne, e poi le righe:



Al di fuori dello standard, esistono numerose metodologie per effettuare questo mapping, e nessuna di esse prevale sulle altre, quindi, in pratica, le scelte di progettazione sono lasciate al “buon senso” del progettista.