

# XPath



Moreno Marzolla  
INFN Sezione di Padova  
moreno.marzolla@pd.infn.it  
<http://www.dsi.unive.it/~marzolla>

## Riferimenti

- Erik T. Ray, *Learning XML*, O'Reilly, First Edition, January 2001 ISBN: 0-59600-046-4, 368 pages
- *XML Path Language (XPath)*, Version 1.0, W3C Recommendation 16 November 1999, <http://www.w3.org/TR/xpath>
- Parte di questi lucidi sono basati sul materiale del corso di Tecnologie Web, prof. Fabio Vitali, Università di Bologna

# XPath

- XPath è un meccanismo (non basato su una sintassi XML) per descrivere locazioni all'interno di documenti XML
- Un XPath è una espressione che restituisce un oggetto di uno di questi quattro tipi:
  - Un booleano
  - Una stringa
  - Un numero
  - Un insieme di nodi (nodi elemento, nodi attributi, nodi testo)

## Alcuni esempi / 1

- **child::para**
  - selects the para element children of the context node
- **child::\***
  - selects all element children of the context node
- **child::text()**
  - selects all text node children of the context node
- **child::node()**
  - selects all the children of the context node, whatever their node type
- **attribute::name**
  - selects the name attribute of the context node

## Alcuni Esempi / 2

- **attribute::\***
  - selects all the attributes of the context node
- **descendant::para**
  - selects the para element descendants of the context node
- **ancestor::div**
  - selects all div ancestors of the context node
- **ancestor-or-self::div**
  - selects the div ancestors of the context node and, if the context node is a div element, the context node as well
- **descendant-or-self::para**
  - selects the para element descendants of the context node and, if the context node is a para element, the context node as well

## Alcuni Esempi / 3

- **self::para**
  - selects the context node if it is a para element, and otherwise selects nothing
- **child::chapter/descendant::para**
  - selects the para element descendants of the chapter element children of the context node
- **child::\* / child::para**
  - selects all para grandchildren of the context node
- **/**
  - selects the document root (which is always the parent of the document element)
- **/descendant::para**
  - selects all the para elements in the same document as the context node

## Alcuni Esempi / 4

- **/descendant::olist/child::item**
  - selects all the item elements that have an olist parent and that are in the same document as the context node
- **child::para[position()=1]**
  - selects the first para child of the context node
- **child::para[position()=last()]**
  - selects the last para child of the context node
- **child::para[position()=last()-1]**
  - selects the last but one para child of the context node
- **child::para[position()>1]**
  - selects all the para children of the context node other than the first para child of the context node

## Alcuni Esempi / 5

- **following-sibling::chapter[position()=1]**
  - selects the next chapter sibling of the context node
- **preceding-sibling::chapter[position()=1]**
  - selects the previous chapter sibling of the context node
- **/descendant::figure[position()=42]**
  - selects the forty-second figure element in the document
- **/child::doc/child::chapter[position()=5]/child::section[position()=2]**
  - selects the second section of the fifth chapter of the doc document element
- **child::para[attribute::type="warning"]**
  - selects all para children of the context node that have a type attribute with value warning

## Alcuni Esempi / 6

- `child::para[attribute::type='warning']`  
`[position()=5]`
  - selects the fifth para child of the context node that has a type attribute with value warning
- `child::para[position()=5]`  
`[attribute::type="warning"]`
  - selects the fifth para child of the context node if that child has a type attribute with value warning
- `child::chapter`  
`[child::title='Introduction']`
  - selects the chapter children of the context node that have one or more title children with string-value equal to Introduction
- `child::chapter[child::title]`
  - selects the chapter children of the context node that have one or more title children

## Alcuni Esempi / 7

- `child::*[self::chapter or self::appendix]`
  - selects the chapter and appendix children of the context node
- `child::*[self::chapter or self::appendix][position()=last()]`
  - selects the last chapter or appendix child of the context node

## Location Path

- Il tipo più importante di XPath è il Location Path.
  - Può essere o assoluto o relativo.
  - Un Location Path assoluto inizia con '/'.
- Un Location Path è composto di una sequenza di passi di locazione (*Location Steps*) separati da '/', e letti da sinistra a destra.
  - Es.: `/child::doc/child::chapter/descendant::para` identifica tutti gli elementi "para" che discendano da ogni elemento "chapter" che sia figlio diretto della radice "doc" del documento XML.

## Location Step

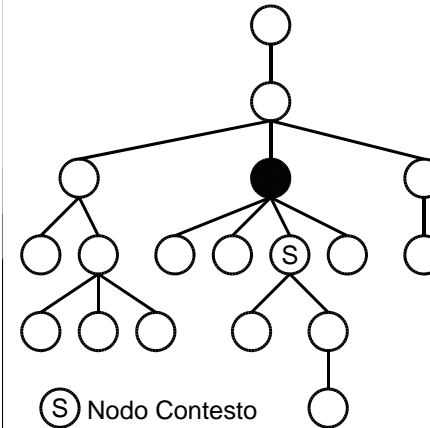
- Un location step ha tre parti:
  - Un **asse**, che individua la direzione di specifica del location step nell'albero
  - Un **test di nodo**, che individua il tipo e il nome completo del nodo identificato dal location step
  - Zero o più **predicati** che raffinano ulteriormente l'insieme di nodi selezionati dal location step
- La sintassi è:
  - `axis::test[pred1][pred2]...[pred N]`

## Assi

- Gli assi identificano la direzione rispetto alla struttura del documento in cui andare a cercare l'oggetto da restituire rispetto al nodo contesto (NC). Tra gli assi possibili troviamo:
  - *child, descendant*: figlio diretto e a qualunque livello del NC
  - *parent, ancestor*: il genitore immediato / qualunque livello del NC
  - *self, namespace*: il NC; il nodo namespace del NC
  - *attribute*: gli attributi del NC
  - *preceding-sibling, following-sibling*: i nodi allo stesso livello ma precedenti o seguenti il NC.
  - *preceding, following*: i nodi a qualunque livello (ma fuori al NC) che precedono o seguono il NC.
  - *descendant-or-self, ancestor-or-self*: come descendant e ancestor, ma considerando anche il NC.

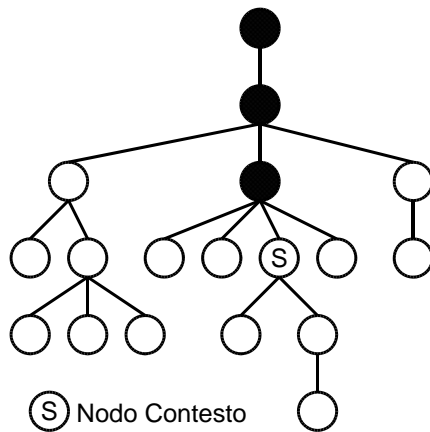
## Esempio

- Asse parent::



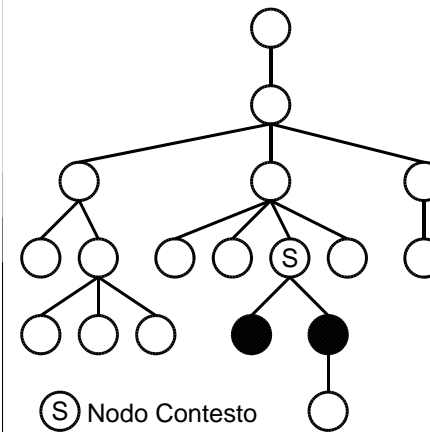
## Esempio

- Asse ancestor::



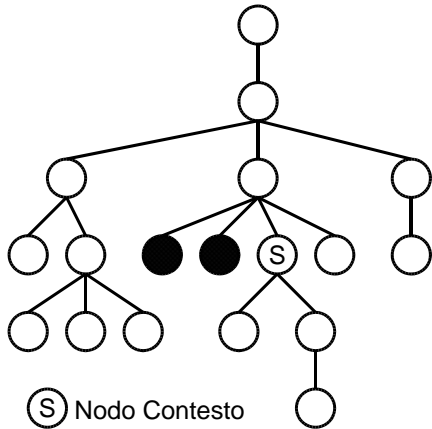
## Esempio

- Asse child::



## Esempio

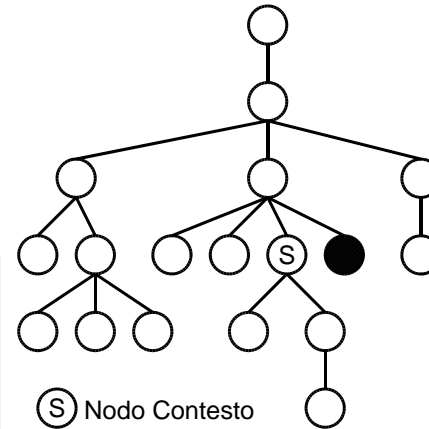
- Asse preceding-sibling::



(S) Nodo Contesto

## Esempio

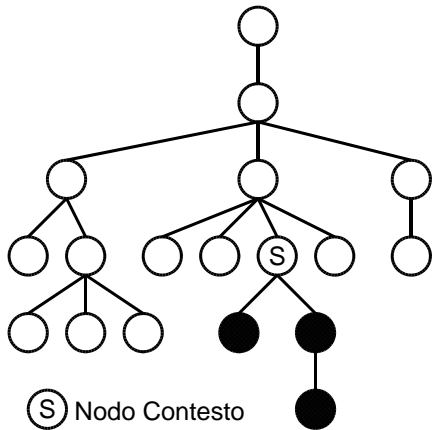
- Asse following-sibling::



(S) Nodo Contesto

## Esempio

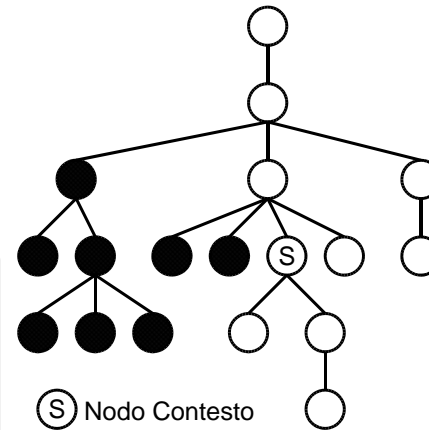
- Asse descendant::



(S) Nodo Contesto

## Esempio

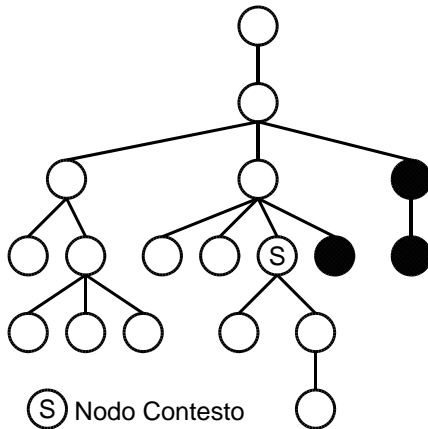
- Asse preceding::



(S) Nodo Contesto

## Esempio

- Asse following::



## Sintassi abbreviata

- In alcuni casi esistono delle forme abbreviate usabili invece della sintassi completa:
  - Child::x si può abbreviare con x
  - Attribute::a si può abbreviare con @a
  - Descendant si può abbreviare con '//', self con '.', parent con '..'
- Esempi:
  - /doc/chapter[5]/section[2]
    - la seconda sezione del quinto capitolo del documento.
  - chapter//para
    - tutti i para discendenti a qualunque livello di ogni nodo chapter figlio del NC
  - //para
    - tutti i para discendenti a qualunque livello della radice del documento.

## Altri esempi di XPath

- `para[@type="warning"]`
  - tutti i para figli del NC che abbiano l'attributo type uguale a "warning".
- `para[@type="warning"][5]`
  - il quinto para figlio di NC ad avere l'attributo type uguale a "warning".
- `para[5][@type="warning"]`
  - il quinto para figlio di NC, ma solo se ha l'attributo type uguale a "warning".
- `chapter[title]`
  - il "chapter" figlio del NC che contenga uno o più elementi "title"

## Un esempio

```
<prenotazioni>
  <biglietto nome="Luca Verdi">
    <voloAndata data="1-8-2005">
      <tappa t="Bologna"/><tappa t="Frankfurt"/><tappa t="NewYork"/>
    </voloAndata>
    <voloRitorno data="8-8-2005">
      <tappa t="NewYork"/><tappa t="Roma"/><tappa t="Bologna"/>
    </voloRitorno>
    <prezzo>990</prezzo>
  </biglietto>
  <biglietto nome="Luigi Bianchi">
    <destinazione>Roma</destinazione>
    <voloAndata data="6-8-2005">
      <tappa t="Bologna"/><tappa t="Roma"/>
    </voloAndata>
    <voloRitorno data="7-8-2005">
      <tappa t="Roma"/><tappa t="Napoli"/>
    </voloRitorno>
    <prezzo>240</prezzo>
  </biglietto>
</-- ... -->
</prenotazioni>
```

1. Tutti i biglietti di Luca Verdi
2. Tutti i biglietti che costano più di 1000 euro
3. Tutte le tappe di tutti i voli
4. Tutti i biglietti del 6 agosto
5. Tutti i biglietti che passano per Roma
6. Le date di tutti i biglietti che arrivano a Napoli

## Un esempio / 1

```
<prenotazioni>
<biglietto nome="Luca Verdi">
  <voloAndata data="1-8-2005">
    <tappa t="Bologna"/><tappa t="Frankfurt"/><tappa t="NewYork"/>
  </voloAndata>
  <voloRitorno data="8-8-2005">
    <tappa t="NewYork"/><tappa t="Roma"/><tappa t="Bologna"/>
  </voloRitorno>
  <prezzo>990</prezzo>
</biglietto>
<biglietto nome="Luigi Bianchi">
  <destinazione>Roma</destinazione>
  <voloAndata data="6-8-2005">
    <tappa t="Bologna"/><tappa t="Roma"/>
  </voloAndata>
  <voloRitorno data="7-8-2005">
    <tappa t="Roma"/><tappa t="Bologna"/>
  </voloRitorno>
  <prezzo>240</prezzo>
</biglietto>
<!-- ... -->
</prenotazioni>
```

Tutti i biglietti di Luca Verdi

//prenotazioni/biglietto[@nome="Luca Verdi"]

## Un esempio / 2

```
<prenotazioni>
<biglietto nome="Luca Verdi">
  <voloAndata data="1-8-2005">
    <tappa t="Bologna"/><tappa t="Frankfurt"/><tappa t="NewYork"/>
  </voloAndata>
  <voloRitorno data="8-8-2005">
    <tappa t="NewYork"/><tappa t="Roma"/><tappa t="Bologna"/>
  </voloRitorno>
  <prezzo>990</prezzo>
</biglietto>
<biglietto nome="Luigi Bianchi">
  <destinazione>Roma</destinazione>
  <voloAndata data="6-8-2005">
    <tappa t="Bologna"/><tappa t="Roma"/>
  </voloAndata>
  <voloRitorno data="7-8-2005">
    <tappa t="Roma"/><tappa t="Bologna"/>
  </voloRitorno>
  <prezzo>240</prezzo>
</biglietto>
<!-- ... -->
</prenotazioni>
```

Tutti i biglietto che costano piu' di 1000 euro

//prenotazioni/biglietto[prezzo > 1000]

## Un esempio / 3

```
<prenotazioni>
<biglietto nome="Luca Verdi">
  <voloAndata data="1-8-2005">
    <tappa t="Bologna"/><tappa t="Frankfurt"/><tappa t="NewYork"/>
  </voloAndata>
  <voloRitorno data="8-8-2005">
    <tappa t="NewYork"/><tappa t="Roma"/><tappa t="Bologna"/>
  </voloRitorno>
  <prezzo>990</prezzo>
</biglietto>
<biglietto nome="Luigi Bianchi">
  <destinazione>Roma</destinazione>
  <voloAndata data="6-8-2005">
    <tappa t="Bologna"/><tappa t="Roma"/>
  </voloAndata>
  <voloRitorno data="7-8-2005">
    <tappa t="Roma"/><tappa t="Bologna"/>
  </voloRitorno>
  <prezzo>240</prezzo>
</biglietto>
<!-- ... -->
</prenotazioni>
```

Tutte le tappe di tutti i voli

//tappa

## Un esempio / 4

```
<prenotazioni>
<biglietto nome="Luca Verdi">
  <voloAndata data="1-8-2005">
    <tappa t="Bologna"/><tappa t="Frankfurt"/><tappa t="NewYork"/>
  </voloAndata>
  <voloRitorno data="8-8-2005">
    <tappa t="NewYork"/><tappa t="Roma"/><tappa t="Bologna"/>
  </voloRitorno>
  <prezzo>990</prezzo>
</biglietto>
<biglietto nome="Luigi Bianchi">
  <destinazione>Roma</destinazione>
  <voloAndata data="6-8-2005">
    <tappa t="Bologna"/><tappa t="Roma"/>
  </voloAndata>
  <voloRitorno data="7-8-2005">
    <tappa t="Roma"/><tappa t="Bologna"/>
  </voloRitorno>
  <prezzo>240</prezzo>
</biglietto>
<!-- ... -->
</prenotazioni>
```

Tutti i biglietti del 6 agosto 2005

//biglietto[\*/@data="6-8-2005"]

## Un esempio / 5

```
<prenotazioni>
<biglietto nome="Luca Verdi">
<voloAndata data="1-8-2005">
  <tappa t="Bologna"/><tappa t="Frankfurt"/><tappa t="NewYork"/>
</voloAndata>
<voloRitorno data="8-8-2005">
  <tappa t="NewYork"/><tappa t="Roma"/><tappa t="Bologna"/>
</voloRitorno>
<prezzo>990</prezzo>
</biglietto>
<biglietto nome="Luigi Bianchi">
<destinazione>Roma</destinazione>
<voloAndata data="6-8-2005">
  <tappa t="Bologna"/><tappa t="Roma"/>
</voloAndata>
<voloRitorno data="7-8-2005">
  <tappa t="Roma"/><tappa t="Bologna"/>
</voloRitorno>
<prezzo>240</prezzo>
</biglietto>
<!-- ... -->
</prenotazioni>
```

Tutti i biglietti che passano per Roma

```
//biglietto[//tappa[@t="Roma"]
```

## Un esempio / 6

```
<prenotazioni>
<biglietto nome="Luca Verdi">
<voloAndata data="1-8-2005">
  <tappa t="Bologna"/><tappa t="Frankfurt"/><tappa t="NewYork"/>
</voloAndata>
<voloRitorno data="8-8-2005">
  <tappa t="NewYork"/><tappa t="Roma"/><tappa t="Bologna"/>
</voloRitorno>
<prezzo>990</prezzo>
</biglietto>
<biglietto nome="Luigi Bianchi">
<destinazione>Roma</destinazione>
<voloAndata data="6-8-2005">
  <tappa t="Bologna"/><tappa t="Roma"/>
</voloAndata>
<voloRitorno data="7-8-2005">
  <tappa t="Roma"/><tappa t="Bologna"/>
</voloRitorno>
<prezzo>240</prezzo>
</biglietto>
<!-- ... -->
</prenotazioni>
```

Le date di tutti i biglietti che arrivano a Napoli

```
/prenotazioni/biglietto/*[./tappa[last()]/@t="Napoli"]/@data
```