

Introducción a XPath



Jorge Castellanos Vega

Introducción

- XPath → Lenguaje que nos permite navegar a través de los elementos y atributos que forman un documento XML
 - Para ello utiliza expresiones basadas en las que se utilizan para navegar por el árbol de directorios y ficheros de un sistema operativo
- XPath Significa XML Path Language
- Contiene aproximadamente 200 funciones para ayudarnos en las operaciones con los nodos:
 - Funciones de cadena
 - Funciones numéricas
 - Funciones lógicas
 - Funciones de fecha y hora
 - ...

Introducción

- Varias versiones
 - XPath 1.0 → Noviembre 1999
 - XPath 2.0 → Enero 2007
 - XPath 3.0 → Abril 2014
 - Xpath 3.1 → Marzo 2017

Nodos XPath

- En Xpath hay siete tipos de nodos:
 - **Elementos**. Cualquier nodo que no es final
 - **Atributos**. Atributos de los elementos.
 - **Texto**. Texto contenido en las etiquetas
 - **Espacio de nombres**
 - **Instrucciones de proceso**
 - **Comentarios**
 - **Nodos de documento**

Nodos XPath

- Los **nodos atributo** y de **texto** no pueden tener descendientes
- El nodo atributo no tiene realmente la consideración de hijo, es simplemente un elemento contenido en la etiqueta
- El texto de una etiqueta si se considera hijo

Nodos XPath

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <MiBibliotecaMP3>
3   <archivo>
4     <canCIÓN>Hangar 18</canCIÓN>
5     <artista cantante="Dave Mustaine">Megadeth</artista>
6     <disco discográfica="capitol" año="1990">Rust in Peace</disco>
7     <puntuación>9</puntuación>
8   </archivo>
9 </MiBibliotecaMP3>
```

- **Nodos:**
 - Nodo raíz → <MiBibliotecaMP3>
 - Nodo de elemento → <artista>
 - Nodo de atributo → año="1990"
- Valores atómicos:
 - Nodos sin padre o sin hijos
 - P.e: "1990", Hangar 18

Nodos XPath

- **Relaciones entre nodos**

- **Padre**

- Cada elemento y atributo tiene un padre
 - En el ejemplo, **archivo** es el padre de **canción**, **artista**, **disco** y **puntuación**

- **Hijo**

- Cada nodo puede tener uno o más hijos
 - En el ejemplo, **archivo** es hijo de **MiBibliotecaMP3**

- **Hermanos**

- Nodos que comparten padre

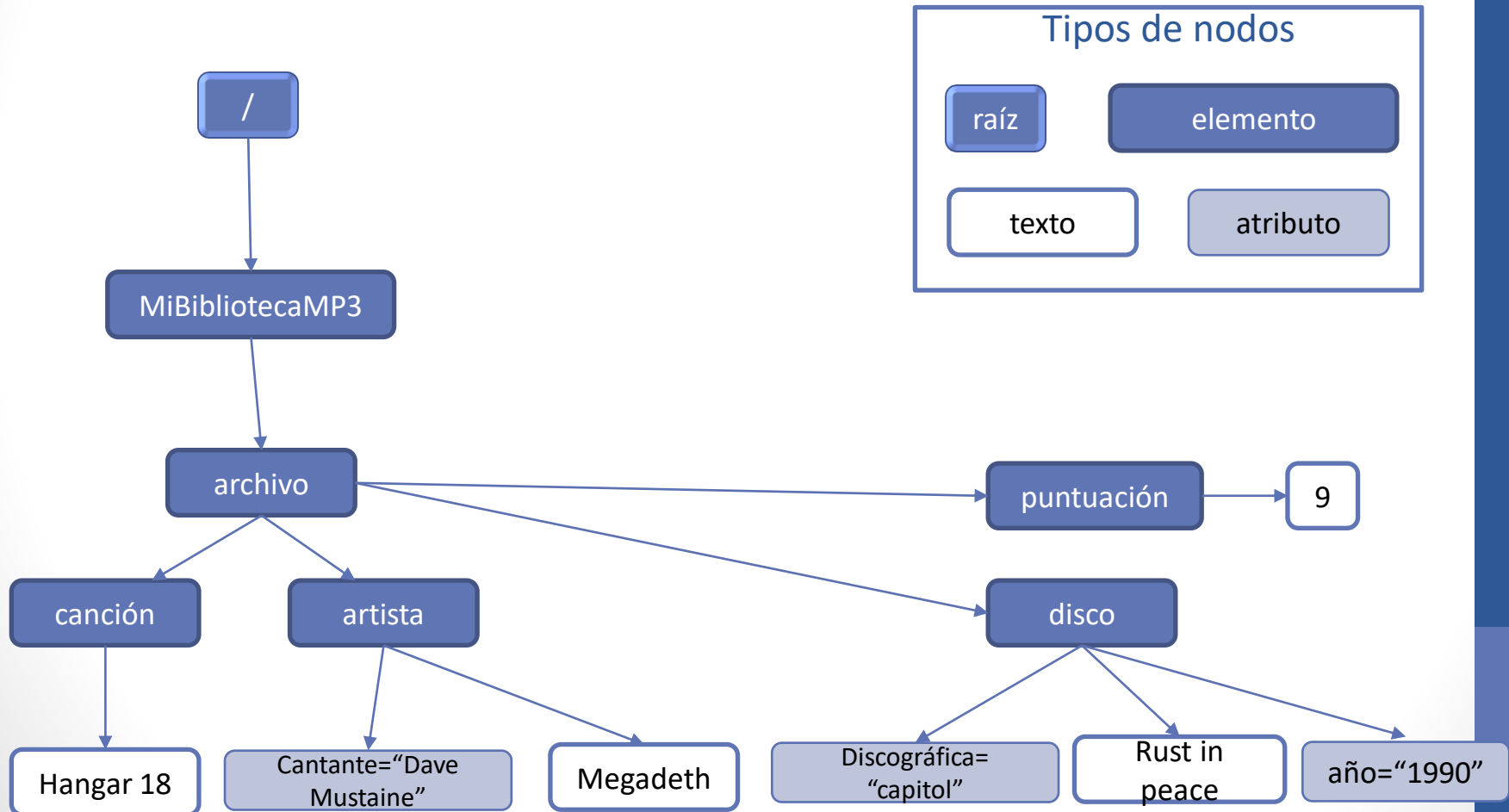
Nodos XPath

- **Antepasados**
 - Todos los nodos de los que desciende un nodo
- **Descendientes**
 - Todos los nodos que descienden de un nodo determinado

Nodos XPath

- Jerarquía

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <MiBibliotecaMP3>
3   <archivo>
4     <canción>Hangar 18</canción>
5     <artista cantante="Dave Mustaine">Megadeth</artista>
6     <disco discográfica="capitol" año="1990">Rust in Peace</disco>
7     <puntuación>9</puntuación>
8   </archivo>
9 </MiBibliotecaMP3>
```



Sintaxis

- Una expresión **XPath** representa un recorrido en el árbol del documento.
 - Parecida a las expresiones de recorrer un árbol de directorios y archivos de un disco
- Devuelve los nodos que se ajustan al recorrido indicado en la expresión **XPath**

Sintaxis

- **Seleccionando nodos**

| Expresión | Acción |
|-----------------|--|
| Nombre del nodo | Selecciona todos los nodos con ese nombre |
| / | Selecciona desde el nodo raíz |
| // | Selecciona todos los descendientes del nodo actual que coincidan con la selección |
| . | Selecciona el nodo actual |
| .. | Selecciona el padre del nodo actual |
| @ | Selecciona atributos |

Sintaxis - Ejemplos

- Supongamos el siguiente documento XML

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <MiBibliotecaMP3>
3   <archivo>
4     <canCIÓN>Hangar 18</canCIÓN>
5     <artista cantante="Dave Mustaine">Megadeth</artista>
6     <disco discográfica="capitol" año="1990">Rust in Peace</disco>
7     <puntuación>9</puntuación>
8   </archivo>
9   <archivo>
10    <canCIÓN>Peace Sells</canCIÓN>
11    <artista cantante="Dave Mustaine">Megadeth</artista>
12    <disco discográfica="capitol" año="1986">Peace Sells...But Who's Buying</disco>
13    <puntuación>9</puntuación>
14  </archivo>
15  <archivo>
16    <canCIÓN>Master of Puppets</canCIÓN>
17    <artista cantante="James Hetfield">Metallica</artista>
18    <disco discográfica="Elektra" año="1986">Master of Puppets</disco>
19    <puntuación>10</puntuación>
20  </archivo>
21  <archivo>
22    <canCIÓN>Among The Living</canCIÓN>
23    <artista cantante="Joey Belladonna">Anthrax</artista>
24    <disco discográfica="Megaforce" año="1987">Among The Living</disco>
25    <puntuación>8</puntuación>
26  </archivo>
27 </MiBibliotecaMP3>
```

Sintaxis - Ejemplos

- Expresión → /MiBibliotecaMP3

Resultado

Equivalente a → /

```
1 <MiBibliotecaMP3>
2 <archivo>
3   <canCIÓN>Hangar 18</canCIÓN>
4   <artista cantante="Dave Mustaine">Megadeth</artista>
5   <disco discográfica="capitol" año="1990">Rust in Peace</disco>
6   <puntuación>9</puntuación>
7 </archivo>
8 <archivo>
9   <canCIÓN>Peace Sells</canCIÓN>
10  <artista cantante="Dave Mustaine">Megadeth</artista>
11  <disco discográfica="capitol" año="1986">Peace Sells...But Who's Buying</disco>
12  <puntuación>9</puntuación>
13 </archivo>
14 <archivo>
15  <canCIÓN>Master of Puppets</canCIÓN>
16  <artista cantante="James Hetfield">Metallica</artista>
17  <disco discográfica="Elektra" año="1986">Master of Puppets</disco>
18  <puntuación>10</puntuación>
19 </archivo>
20 <archivo>
21  <canCIÓN>Among The Living</canCIÓN>
22  <artista cantante="Joey Belladonna">Anthrax</artista>
23  <disco discográfica="Megaforce" año="1987">Among The Living</disco>
24  <puntuación>8</puntuación>
25 </archivo>
26 </MiBibliotecaMP3>
```

Sintaxis - Ejemplos

- Expresión → /MiBibliotecaMP3/archivo

Resultado

```
1 <archivo>
2   <canción>Hangar 18</canción>
3   <artista cantante="Dave Mustaine">Megadeth</artista>
4   <disco discográfica="capitol" año="1990">Rust in Peace</disco>
5   <puntuación>9</puntuación>
6 </archivo>
7 <archivo>
8   <canción>Peace Sells</canción>
9   <artista cantante="Dave Mustaine">Megadeth</artista>
10  <disco discográfica="capitol" año="1986">Peace Sells...But Who's Buying</disco>
11  <puntuación>9</puntuación>
12 </archivo>
13 <archivo>
14  <canción>Master of Puppets</canción>
15  <artista cantante="James Hetfield">Metallica</artista>
16  <disco discográfica="Elektra" año="1986">Master of Puppets</disco>
17  <puntuación>10</puntuación>
18 </archivo>
19 <archivo>
20  <canción>Among The Living</canción>
21  <artista cantante="Joey Belladonna">Anthrax</artista>
22  <disco discográfica="Megaforce" año="1987">Among The Living</disco>
23  <puntuación>8</puntuación>
24 </archivo>
25
```

Sintaxis - Ejemplos

- **Expresión** → /MiBibliotecaMP3/archivo/canción

Resultado

```
1 <canción>Hangar 18</canción>  
2 <canción>Peace Sells</canción>  
3 <canción>Master of Puppets</canción>  
4 <canción>Among The Living</canción>
```

- **Expresión** → /MiBibliotecaMP3/archivo/disco

Resultado

```
1 <disco discográfica="capitol" año="1990">Rust in Peace</disco>  
2 <disco discográfica="capitol" año="1986">Peace Sells...But Who's Buying</disco>  
3 <disco discográfica="Elektra" año="1986">Master of Puppets</disco>  
4 <disco discográfica="Megaforce" año="1987">Among The Living</disco>
```

Sintaxis - Ejemplos

- **Expresión** → /MiBibliotecaMP3/archivo/disco/@discográfica

Resultado

| | |
|---|--------------------------|
| 1 | discográfica="capitol" |
| 2 | discográfica="capitol" |
| 3 | discográfica="Elektra" |
| 4 | discográfica="Megaforce" |

- **Expresión** → /MiBibliotecaMP3/archivo/artista/@cantante

Resultado

| |
|----------------------------|
| cantante="Dave Mustaine" |
| cantante="Dave Mustaine" |
| cantante="James Hetfield" |
| cantante="Joey Belladonna" |

Sintaxis - Ejemplos

- **Expresión** → /MiBibliotecaMP3/archivo/artista/..

Resultado

```
<archivo>
  <canción>Hangar 18</canción>
  <artista cantante="Dave Mustaine">Megadeth</artista>
  <disco discográfica="capitol" año="1990">Rust in Peace</disco>
  <puntuación>9</puntuación>
</archivo>
<archivo>
  <canción>Peace Sells</canción>
  <artista cantante="Dave Mustaine">Megadeth</artista>
  <disco discográfica="capitol" año="1986">Peace Sells...But Who's Buying</disco>
  <puntuación>9</puntuación>
</archivo>
<archivo>
  <canción>Master of Puppets</canción>
  <artista cantante="James Hetfield">Metallica</artista>
  <disco discográfica="Elektra" año="1986">Master of Puppets</disco>
  <puntuación>10</puntuación>
</archivo>
<archivo>
  <canción>Among The Living</canción>
  <artista cantante="Joey Belladonna">Anthrax</artista>
  <disco discográfica="Megaforce" año="1987">Among The Living</disco>
  <puntuación>8</puntuación>
</archivo>
```

Sintaxis - Ejemplos

- Expresión → //archivo

Resultado

```
1 <archivo>
2   <canción>Hangar 18</canción>
3   <artista cantante="Dave Mustaine">Megadeth</artista>
4   <disco discográfica="capitol" año="1990">Rust in Peace</disco>
5   <puntuación>9</puntuación>
6 </archivo>
7 <archivo>
8   <canción>Peace Sells</canción>
9   <artista cantante="Dave Mustaine">Megadeth</artista>
10  <disco discográfica="capitol" año="1986">Peace Sells...But Who's Buying</disco>
11  <puntuación>9</puntuación>
12 </archivo>
13 <archivo>
14  <canción>Master of Puppets</canción>
15  <artista cantante="James Hetfield">Metallica</artista>
16  <disco discográfica="Elektra" año="1986">Master of Puppets</disco>
17  <puntuación>10</puntuación>
18 </archivo>
19 <archivo>
20  <canción>Among The Living</canción>
21  <artista cantante="Joey Belladonna">Anthrax</artista>
22  <disco discográfica="Megaforce" año="1987">Among The Living</disco>
23  <puntuación>8</puntuación>
24 </archivo>
```

Sintaxis - Ejemplos

- **Expresión** → //canción

Resultado

```
1 <canción>Hangar 18</canción>  
2 <canción>Peace Sells</canción>  
3 <canción>Master of Puppets</canción>  
4 <canción>Among The Living</canción>  
5
```

- **Expresión** → //disco/./artista/@cantante

Resultado

```
cantante="Dave Mustaine"  
cantante="Dave Mustaine"  
cantante="James Hetfield"  
cantante="Joey Belladonna"
```

Sintaxis - Ejemplos

- **Expresión** → `//@discográfica`

Resultado

| | |
|---|---------------------------------------|
| 1 | <code>discográfica="capitol"</code> |
| 2 | <code>discográfica="capitol"</code> |
| 3 | <code>discográfica="Elektra"</code> |
| 4 | <code>discográfica="Megaforce"</code> |

- **Expresión** → `//@discográfica/../../canción`

Resultado

| | |
|---|---|
| 1 | <code><canción>Hangar 18</canción></code> |
| 2 | <code><canción>Peace Sells</canción></code> |
| 3 | <code><canción>Master of Puppets</canción></code> |
| 4 | <code><canción>Among The Living</canción></code> |

Sintaxis

- Es posible **seleccionar varias rutas** con el operador | (AND)
- **Expresión** → //canción | //artista/@cantante

Resultado

```
1 <canción>Hangar 18</canción>  
2 cantante="Dave Mustaine"  
3 <canción>Peace Sells</canción>  
4 cantante="Dave Mustaine"  
5 <canción>Master of Puppets</canción>  
6 cantante="James Hetfield"  
7 <canción>Among The Living</canción>  
8 cantante="Joey Belladonna"
```

Sintaxis

- **Predicados**

- Se utilizan para seleccionar nodos que contengan un valor determinado
- Siempre se expresan escritos entre corchetes
- Las expresiones utilizadas pueden contener operadores:
- Aritméticos (+, -, *, /, div, mod)
- Comparación (=, !=, >, <, >=, <=)
- Lógicos (and, or, not)
- Es posible combinar varios predicados uno a continuación de otro.
 - Cada uno se aplica sobre el resultado del anterior

Sintaxis

- **Predicados**

| Predicado | Acción | Ejemplo |
|--------------|--|--|
| [n] | Selecciona el elementos n-ésimo El primer elemento se identifica por el 1 | //canción[1] |
| [last()] | Selecciona el último elemento que es hijo de los nodos seleccionados | //canción[last()] //canción[last()-1] |
| [position()] | Permite seleccionar elementos a partir de la posición en la que se encuentren en los resultados | //canción[position()>2] |
| [@atrib] | Devuelve los nodos que tienen el atributo atrib o el valor del atributo si se indica en la expresión | //disco[@año] //disco[@año="1986"] |

Sintaxis – Ejemplos predicados

- **Expresión** → `//disco[@año="1990" or @año="1987"]`

```
<disco discográfica="capitol" año="1990">Rust in Peace</disco>  
<disco discográfica="Megaforce" año="1987">Among The Living</disco>
```

- **Expresión** → `//disco[not(@año="1990")]`

```
1 <disco discográfica="capitol" año="1986">Peace Sells...But Who's Buying</disco>  
2 <disco discográfica="Elektra" año="1986">Master of Puppets</disco>  
3 <disco discográfica="Megaforce" año="1987">Among The Living</disco>
```

- **Expresión** → `//disco[not(@año="1990")][last()]`

```
<disco discográfica="Megaforce" año="1987">Among The Living</disco>
```


Sintaxis – Ejemplos predicados

- **Expresión** → //archivo[puntuacion=8]

```
<archivo>  
  <canción>Among The Living</canción>  
  <artista cantante="Joey Belladonna">Anthrax</artista>  
  <disco discográfica="Megaforce" año="1987">Among The Living</disco>  
  <puntuacion>8</puntuacion>  
</archivo>
```

- **Expresión** → //archivo[puntuacion=8]/artista

```
<artista cantante="Joey Belladonna">Anthrax</artista>
```

- **Expresión** → //archivo[puntuacion=8]/artista/@cantante

```
cantante="Joey Belladonna"
```

Sintaxis – Ejemplos predicados

- **Expresión** → //archivo[2]/disco/@discográfica

- **Resultado:** discográfica="capitol"

- **Expresión** → //disco[position()>=3]/@discográfica

| | |
|---|-------------------------------|
| 1 | <u>discográfica="Elektra"</u> |
| 2 | discográfica="Megaforce" |

- **Expresión**

→ //disco[position()>=2 and position()<=last()]/@discográfica

| |
|-------------------------------|
| <u>discográfica="capitol"</u> |
| discográfica="Elektra" |
| discográfica="Megaforce" |

Sintaxis – Texto del elemento

- Obteniendo el contenido de un elemento – **text()**

- **Expresión** → `//artista/text()`

| |
|-----------|
| Megadeth |
| Megadeth |
| Metallica |
| Anthrax |

- **Expresión** → `//disco[@discográfica="Elektra"]/text()`

- Resultado: Master of Puppets

- **Expresión** → `//disco[@discográfica="Elektra"]/../artista/text()`

- Resultado: Metallica

Sintaxis - Comodines

- **Comodines**

- Es posible utilizar comodines para seleccionar nodos de los que no conocemos el nombre

| * | Comodin para elementos |
|--------|--------------------------------------|
| @* | Comodín para atributos |
| node() | Comodín para nodos de cualquier tipo |

- **Expresión** → //archivo/disco/@*

| | |
|---|-------------------------------|
| 1 | <u>discográfica="capitol"</u> |
| 2 | año="1990" |
| 3 | discográfica="capitol" |
| 4 | año="1986" |
| 5 | discográfica="Elektra" |
| 6 | año="1986" |
| 7 | discográfica="Megaforce" |
| 8 | año="1987" |

Sintaxis – Ejemplos comodines

- **Expresión**

- `//archivo/*` (en este caso equivale a `//archivo/node()`)

```
1 <canción>Hangar 18</canción>
2 <artista cantante="Dave Mustaine">Megadeth</artista>
3 <disco discográfica="capitol" año="1990">Rust in Peace</disco>
4 <puntuación>9</puntuación>
5 <canción>Peace Sells</canción>
6 <artista cantante="Dave Mustaine">Megadeth</artista>
7 <disco discográfica="capitol" año="1986">Peace Sells...But Who's Buying</disco>
8 <puntuación>9</puntuación>
9 <canción>Master of Puppets</canción>
10 <artista cantante="James Hetfield">Metallica</artista>
11 <disco discográfica="Elektra" año="1986">Master of Puppets</disco>
12 <puntuacion>10</puntuacion>
13 <canción>Among The Living</canción>
14 <artista cantante="Joey Belladonna">Anthrax</artista>
15 <disco discográfica="Megaforce" año="1987">Among The Living</disco>
16 <puntuacion>8</puntuacion>
```

Referencias

- W3School - <https://www.w3schools.com>