

XPath: XML Path language

- [Qué es XPath](#)
- [Árbol del documento](#)
 - [Tipos de nodos](#)
- [Sintaxis de la expresiones XPath](#)
- [Sintaxis abreviada](#)
 - [Ejes](#)
 - [Nodos de comprobación](#)
 - [Predicados](#)
 - [Expresiones anidadas](#)

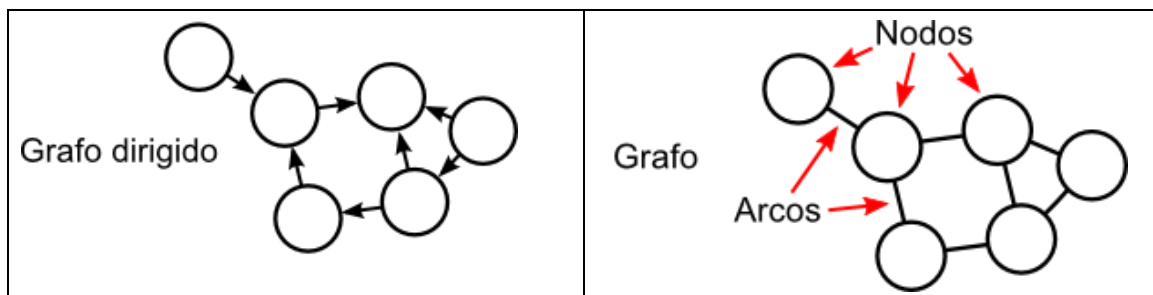
Qué es XPath

XPath es un lenguaje que permite seleccionar nodos de un documento XML y calcular valores a partir de su contenido. Existen dos versiones de XPath aprobadas por el W3C.

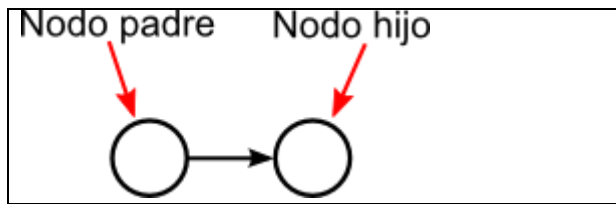
- noviembre de 1999: [XML Path Language 1.0](#)
- enero de 2007: [XML Path Language 2.0](#)
- diciembre de 2010: [XML Path Language 2.0 \(2º edición\)](#)
- abril de 2014 Xpath 3.0 y marzo 2017 Xpath 3.1

Árbol del documento

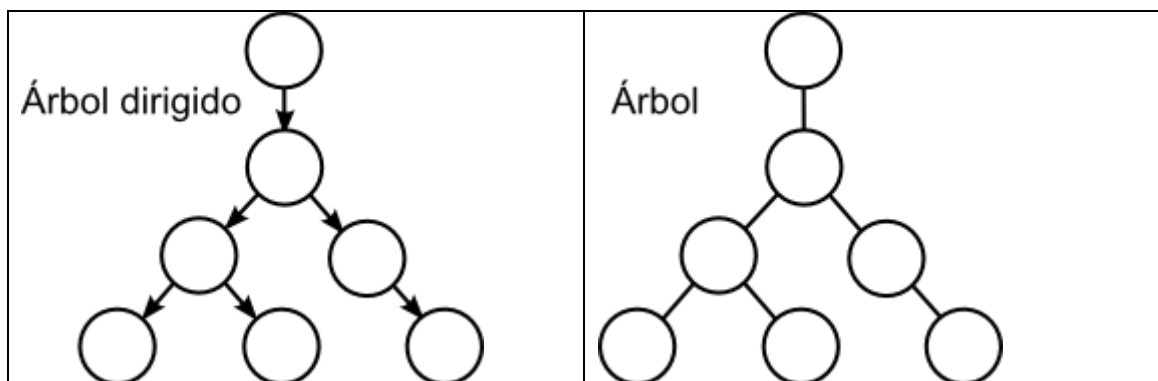
XPath considera un documento XML como un árbol de nodos. En Informática, un árbol es una estructura de datos que equivale a un árbol matemático. En Matemáticas un árbol es un caso particular de grafo. Los siguientes términos definidos en teoría de grafos se utilizan también en Informática y en XPath:



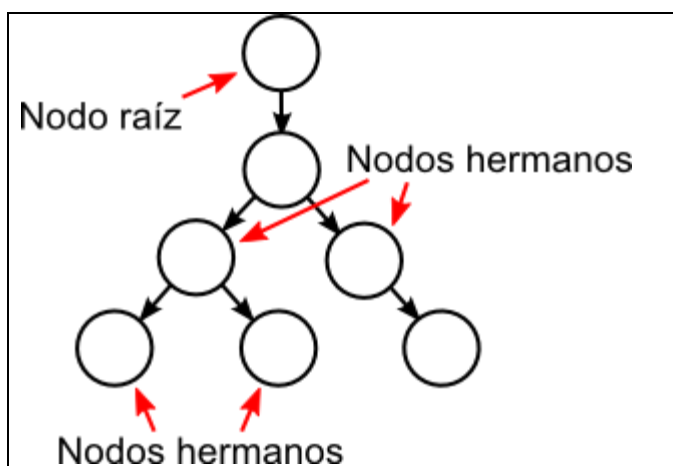
- Un **grafo** es un conjunto de objetos llamados nodos o vértices unidos por enlaces llamados arcos o aristas. Un **grafo dirigido** es un grafo en el que los arcos tienen dirección.



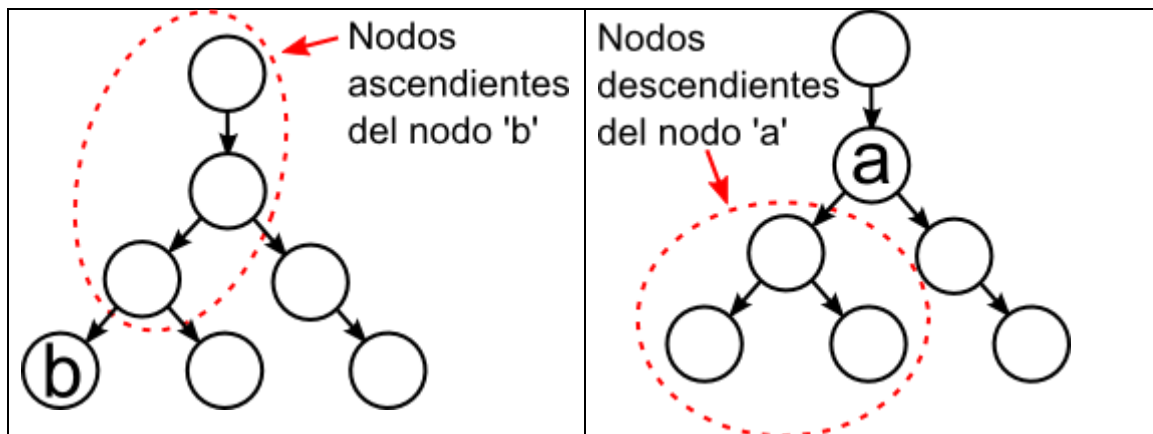
- Cuando dos nodos están unidos por un arco con dirección, el **nodo padre** es el nodo del que parte el arco y el **nodo hijo** es el nodo al que llega el arco.



- Un **árbol** es un grafo en el que cualquier pareja de vértices están conectada por un único camino (es decir, que no hay ciclos). Un **árbol dirigido** es un árbol en el que las aristas tienen dirección y todos los nodos menos uno tienen un único padre.



- El **nodo raíz** de un árbol dirigido es el único nodo sin padre. Los **nodos hermanos** son los nodos que tienen el mismo padre.



- Los **nodos descendientes** de un nodo son todos los nodos a los que se llega desde el nodo: los hijos, los hijos de los hijos, etc. Los **nodos ascendentes** de un nodo son todos los nodos de los que un nodo es descendiente: el padre, el padre del padre, etc.

Tipos de nodos

Un documento XML puede representarse como un árbol dirigido, considerando por ejemplo los elementos como nodos y que un elemento es padre de los elementos que contiene. Pero en XPath no sólo los elementos son nodos, en realidad hay siete tipos de nodos:

- Raíz
- Elemento
- Atributo
- Texto
- Comentario
- Instrucción de procesamiento
- Espacio de nombres

Nota: La declaración DOCTYPE no se considera como nodo.

Por ejemplo, el documento XML siguiente:

```
<?xml version="1.0" encoding="UTF-8"?>
<biblioteca>
  <libro>
    <titulo>La vida está en otra parte</titulo>
    <autor>Milan Kundera</autor>
    <fechaPublicacion año="1973"/>
  </libro>
  <libro>
    <titulo>Pantaleón y las visitadoras</titulo>
    <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor>
    <fechaPublicacion año="1973"/>
  </libro>
  <libro>
    <titulo>Conversación en la catedral</titulo>
    <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor>
    <fechaPublicacion año="1969"/>
  </libro>
</biblioteca>
```

```
</libro>  
</biblioteca>
```

se puede representar mediante el siguiente grafo:



Los nodos atributos y de texto no son como los nodos elemento. Por ejemplo, los nodos atributo y de texto no pueden tener descendientes. En realidad el nodo atributo ni siquiera se considera como hijo, sino como una etiqueta adosada al elemento. El texto contenido por una etiqueta sí que se considera hijo del elemento, aunque las expresiones XPath suelen trabajar con nodos elementos y para referirse a los atributos o al texto se utilizan notaciones especiales.

Sintaxis de la expresiones XPath

Una expresión XPath es una cadena de texto que representa un recorrido en el árbol del documento. Las expresiones más simples se parecen a las rutas de los archivos en el explorador de Windows o en la shell de GNU/Linux.

Evaluar una expresión XPath es buscar si hay nodos en el documento que se ajustan al recorrido definido en la expresión. El resultado de la evaluación son todos los nodos que se ajustan a la expresión. Para poder evaluar una expresión XPath, el documento debe estar bien formado.

Las expresiones XPath se pueden escribir de dos formas distintas:

- [sintaxis abreviada](#): más compacta y fácil de leer
- [sintaxis completa](#): más larga pero con más opciones disponibles

Las expresiones XPath se pueden dividir en pasos de búsqueda. Cada paso de búsqueda se puede a su vez dividir en tres partes:

- eje: indica el nodo o los nodos en los que se realiza la búsqueda
- nodo de comprobación: especifica el nodo o los nodos seleccionados dentro del eje
- predicado: permite restringir los nodos de comprobación

Sintaxis abreviada

Veamos unos ejemplos de expresiones XPath de sintaxis abreviada y el resultado de su evaluación en el documento de ejemplo anterior:

```
<?xml version="1.0" encoding="UTF-8"?>
<biblioteca>
  <libro>
    <titulo>La vida está en otra parte</titulo>
    <autor>Milan Kundera</autor>
    <fechaPublicacion año="1973"/>
  </libro>
  <libro>
    <titulo>Pantaleón y las visitadoras</titulo>
    <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor>
    <fechaPublicacion año="1973"/>
  </libro>
  <libro>
    <titulo>Conversación en la catedral</titulo>
    <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor>
    <fechaPublicacion año="1969"/>
  </libro>
</biblioteca>
```

Ejes

- **/**: si está al principio de la expresión, indica el nodo raíz, si no, indica "hijo".

<code>/biblioteca/libro/autor</code>	<pre><autor>Milan Kundera</autor> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor></pre>
<code>/autor</code>	No devuelve nada porque "autor" no es hijo del nodo raíz.
<code>/biblioteca/autor</code>	No devuelve nada porque "autor" no es hijo de "biblioteca".

<code>/biblioteca//autor</code>	<pre><autor>Milan Kundera</autor> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <autor fechaNacimiento="28/03/1936">Mario</pre>
---------------------------------	--

	Vargas Llosa</autor>
//autor	<autor>Milan Kundera</autor> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor>
//autor//libro	No devuelve nada porque "libro" no es descendiente de "autor".

- //: indica "descendiente" (hijos, hijos de hijos, etc.).
- @atributo: selecciona el atributo.

/biblioteca/libro/autor/@fechaNacimiento	fechaNacimiento="28/03/1936"
/biblioteca/libro/autor[@fechaNacimiento]	fechaNacimiento="28/03/1936"
/biblioteca/libro/@fechaNacimiento	No devuelve nada porque "libro" no tiene el atributo fechaNacimiento.
/biblioteca/libro[@fechaNacimiento]	

- ..: selecciona el elemento padre.

/biblioteca/libro/autor/@fechaNacimiento/..	<autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor>
/biblioteca/libro/autor[@fechaNacimiento]/..	

- Nota: En este ejemplo se seleccionan únicamente los nodos "autor" que tienen el atributo fechaNacimiento.
- |: permite elegir varios recorridos.

//autor//titulo	<titulo>La vida está en otra parte</titulo> <autor>Milan Kundera</autor> <titulo>Pantaleón y las visitadoras</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <titulo>Conversación en la catedral</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor>
-----------------	---

Nodos de comprobación

- node(): selecciona todos los nodos (elementos y texto).

//libro/node()	<titulo>La vida está en otra parte</titulo> <autor>Milan Kundera</autor> <fechaPublicacion año="1973"/> <titulo>Pantaleón y las visitadoras</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1973"/> <titulo>Conversación en la catedral</titulo> <autor fechaNacimiento="28/03/1936">Mario
----------------	---

	Vargas Llosa</autor> <fechaPublicacion año="1969"/>
//autor/node()	Milan Kundera Mario Vargas Llosa Mario Vargas Llosa
//libro//node()	<titulo>La vida está en otra parte</titulo> La vida está en otra parte <autor>Milan Kundera</autor> Milan Kundera <fechaPublicacion año="1973"/> <titulo>Pantaleón y las visitadoras</titulo> Pantaleón y las visitadoras <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> Mario Vargas Llosa <fechaPublicacion año="1973"/> <titulo>Conversación en la catedral</titulo> Conversación en la catedral <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> Mario Vargas Llosa <fechaPublicacion año="1969"/>

- **text()**: selecciona el contenido del elemento (texto).

//autor/text()	Milan Kundera Mario Vargas Llosa Mario Vargas Llosa
//libro/text()	No devuelve nada porque "libro" no contiene texto.

- *****: selecciona todos los elementos

/biblioteca/*	<libro> <titulo>La vida está en otra parte</titulo> <autor>Milan Kundera</autor> <fechaPublicacion año="1973"/> </libro> <libro> <titulo>Pantaleón y las visitadoras</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1973"/> </libro> <libro> <titulo>Conversación en la catedral</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1969"/> </libro>
//autor/*	No devuelve nada porque "autor" sólo contiene texto.
/biblioteca//*	<libro> <titulo>La vida está en otra parte</titulo> <autor>Milan Kundera</autor> <fechaPublicacion año="1973"/> </libro>

	<pre> <titulo>La vida está en otra parte</titulo> <autor>Milan Kundera</autor> <fechaPublicacion año="1973"/> <libro> <titulo>Pantaleón y las visitadoras</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1973"/> </libro> <titulo>Pantaleón y las visitadoras</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1973"/> <libro> <titulo>Conversación en la catedral</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1969"/> </libro> <titulo>Conversación en la catedral</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1969"/> </pre>
--	--

- **@*:** selecciona todos los atributos

//@* - Not found	<pre> año="1973" fechaNacimiento="28/03/1936" año="1973" fechaNacimiento="28/03/1936" año="1969" </pre>
//libro[@*]	No devuelve nada porque "libro" no tiene atributos.
//autor[@*]	<pre> fechaNacimiento="28/03/1936" fechaNacimiento="28/03/1936" </pre>

Predicados

Los predicados se escriben entre corchetes

- **[@atributo]:** selecciona los elementos que tienen el atributo.

//autor[@fechaNacimiento]	<pre> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> </pre>
---------------------------	--

- **[número]:** si hay varios resultados selecciona uno de ellos por número de orden; **last()** selecciona el último de ellos

//libro[1]	<pre><libro> <titulo>La vida está en otra parte</titulo> <autor>Milan Kundera</autor> <fechaPublicacion año="1973"/> </libro></pre>
//libro[last()]	<pre><libro> <titulo>Conversación en la catedral</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1969"/> </libro></pre>
//libro[last()-1]	<pre><libro> <titulo>Pantaleón y las visitadoras</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1973"/> </libro></pre>

Los predicados permiten definir condiciones sobre los valores de los atributos. En las condiciones se pueden utilizar los operadores siguientes:

- operador de unión (OR lógico): |
- operadores lógicos: and, or, not()
- operadores aritméticos: +, -, *, div, mod
- operadores de comparación: =, !=, <, >, <=, >=

Las comparaciones se pueden hacer entre valores de nodos y atributos o con cadenas de texto o numéricas. En el caso de las cadenas de texto deben estar rodeadas por comillas simples o dobles, en el caso de las cadenas numéricas, las comillas son optativas.

- **[condicion]:** selecciona los nodos que cumplen la condición. La condición puede utilizar el valor de un atributo (utilizando @) o el texto que contiene el elemento (utilizando .)

//fechaPublicacion[@año<1970]	<fechaPublicacion año="1969"/>
//fechaPublicacion[@año<1970]/..	<pre><libro> <titulo>Conversación en la catedral</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1969"/> </libro></pre>
//libro[autor='Mario Vargas Llosa']	<pre><libro> <titulo>Pantaleón y las visitadoras</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1973"/> </libro> <libro> <titulo>Conversación en la catedral</titulo></pre>

	<pre> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1969"/> </libro> </pre>
//autor[.="Mario Vargas Llosa"]/..	<pre> <libro> <titulo>Pantaleón y las visitadoras</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1973"/> </libro> <libro> <titulo>Conversación en la catedral</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1969"/> </libro> </pre>

- Se pueden escribir varios predicados seguidos, teniendo en cuenta que cada uno restringe los resultados del anterior, como si estuvieran encadenados por la operación lógica and.

//libro[autor='Mario Vargas Llosa'][fechaPublicacion/@año="1969"]	<pre> <libro> <titulo>Conversación en la catedral</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1969"/> </libro> </pre>
//libro[autor='Mario Vargas Llosa' and fechaPublicacion/@año="1969"]	<pre> <libro> <titulo>Conversación en la catedral</titulo> <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor> <fechaPublicacion año="1969"/> </libro> </pre>

Expresiones anidadas

Las expresiones XPath pueden anidarse, lo que permite definir expresiones más complicadas. Por ejemplo, en el documento utilizado anteriormente:

```

<?xml version="1.0" encoding="UTF-8"?>
<biblioteca>
  <libro>
    <titulo>La vida está en otra parte</titulo>
    <autor>Milan Kundera</autor>
    <fechaPublicacion año="1973"/>
  </libro>
  <libro>
    <titulo>Pantaleón y las visitadoras</titulo>
    <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor>

```

```

    <fechaPublicacion año="1973"/>
  </libro>
</libro>
  <titulo>Conversación en la catedral</titulo>
  <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor>
  <fechaPublicacion año="1969"/>
</libro>
</biblioteca>

```

Un ejemplo de expresión anidada sería, por ejemplo, obtener los títulos de los libros publicados el mismo año que la novela "La vida está en otra parte". Esta información no está directamente almacenada en el documento, pero se puede obtener la respuesta en dos pasos:

- obtener primero el año en que se publicó la novela "La vida está en otra parte":

//titulo[.="La vida está en otra parte"]/..fechaPublicacion/@año	año="1973"
--	------------

- y obtener después los títulos de los libros publicados en 1973:

//fechaPublicacion[@año=1973]/..titulo	<titulo>La vida está en otra parte</titulo> <titulo>Pantaleón y las visitadoras</titulo>
--	---

Estas dos expresiones se pueden unir en una única expresión, sustituyendo en la segunda expresión el valor 1973 por la primera expresión:

//fechaPublicacion[@año=//titulo[.="La vida está en otra parte"]/..fechaPublicacion/@año]/..titulo	<titulo>La vida está en otra parte</titulo> <titulo>Pantaleón y las visitadoras</titulo>
--	---

Otro ejemplo de expresión anidada sería obtener los títulos de los libros del mismo autor que la novela "Pantaleón y las visitadoras". Como en el ejemplo anterior, la respuesta puede obtenerse en dos pasos:

- obtener primero el autor de la novela "Pantaleón y las visitadoras":

//libro[titulo="Pantaleón y las visitadoras"]/autor/text()	Mario Vargas Llosa
--	--------------------

- y obtener después los títulos de los libros escritos por Mario Vargas Llosa:

//libro[autor="Mario Vargas Llosa"]/titulo	<titulo>Pantaleón y las visitadoras</titulo> <titulo>Conversación en la catedral</titulo>
--	--

Estas dos expresiones se pueden unir en una única expresión, sustituyendo en la segunda expresión el valor "Mario Vargas Llosa" por la primera expresión:

<code>//libro[autor=//libro[título="Pantaleón y las visitadoras"]/autor/text()]/título</code>	<pre><título>Pantaleón y las visitadoras</título> <título>Conversación en la catedral</título></pre>
---	--

Un detalle importante es que no hay que escribir la primera expresión entre comillas.

Se puede omitir el nodo de comprobación `text()` de la segunda expresión y escribir la expresión XPath así:

<code>//libro[autor=//libro[título="Pantaleón y las visitadoras"]/autor]/título</code>	<pre><título>Pantaleón y las visitadoras</título> <título>Conversación en la catedral</título></pre>
--	--

Sintaxis abreviada

Ejes

- **/**: si está al principio de la expresión, indica el nodo raíz, si no, indica "hijo".
- **//**: indica "descendiente" (hijos, hijos de hijos, etc.).
- **@atributo**: selecciona el atributo.
- **..**: selecciona el elemento padre.
- **|**: permite elegir varios recorridos.

Nodos de comprobación

- **node()**: selecciona todos los nodos (elementos y texto).
- **text()**: selecciona el contenido del elemento (texto).
- *****: selecciona todos los elementos
- **@***: selecciona todos los atributos

Predicados

Los predicados se escriben entre corchetes

- **[@atributo]**: selecciona los elementos que tienen el atributo.
- **[número]**: si hay varios resultados selecciona uno de ellos por número de orden; **last()** selecciona el último de ellos

- Los predicados permiten definir condiciones sobre los valores de los atributos. En las condiciones se pueden utilizar los operadores siguientes:
 - operador de unión (OR lógico): |
 - operadores lógicos: and, or, not()
 - operadores aritméticos: +, -, *, div, mod
 - operadores de comparación: =, !=, <, >, <=, >=
- **[condicion]**: selecciona los nodos que cumplen la condición. La condición puede utilizar el valor de un atributo (utilizando @) o el texto que contiene el elemento (utilizando .)

Se pueden escribir varios predicados seguidos, teniendo en cuenta que cada uno restringe los resultados del anterior, como si estuvieran encadenados por la operación lógica and.