XPath: XML Path language

Tabla de contenido

[Árbol del documento 1](#_Toc482825068)

[Tipos de nodos 3](#_Toc482825069)

[Sintaxis de la expresiones XPath 4](#_Toc482825070)

[Sintaxis abreviada 5](#_Toc482825071)

[Eje (en inglés, axis) 5](#_Toc482825072)

[Predicado (en inglés, predicate) 7](#_Toc482825073)

[Selección de nodos (en inglés, node test) 9](#_Toc482825074)

[Pasos de búsqueda consecutivos 10](#_Toc482825075)

[Expresiones anidadas 11](#_Toc482825076)

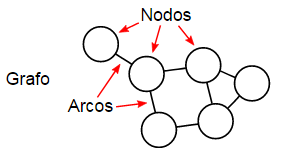
[Ejercicios 13](#_Toc482825077)

[EJERCICIOS: 13](#_Toc482825078)

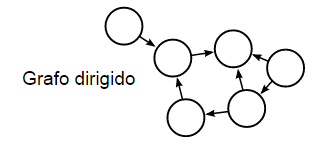
# Árbol del documento

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:

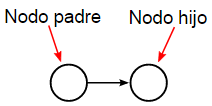
* Un grafo es un conjunto de objetos llamados nodos o vértices unidos por enlaces llamados arcos o aristas.



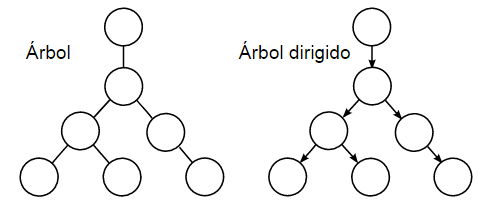
* Un grafo dirigido es un grafo en el que los arcos tienen dirección.



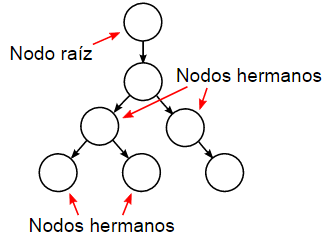
* Cuando dos nodos están unidos por un arco con dirección, el nodo padre es el nodo del que parte el arco y el nodo hijo es el nodo al que llega el arco.



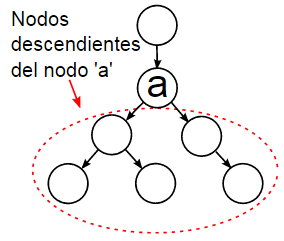
* Un árbol es un grafo en el que cualquier pareja de vértices están conectada por un único camino (es decir, que no hay ciclos). Un árbol dirigido es un árbol en el que las aristas tienen dirección y todos los nodos menos uno tienen un único padre.



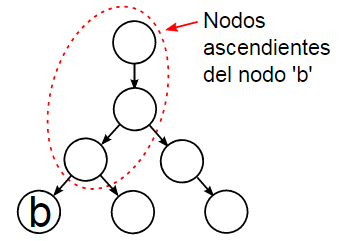
* El nodo raíz de un árbol dirigido es el único nodo sin padre. Los nodos hermanos son los nodos que tienen el mismo padre



* 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 ascendientes 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>

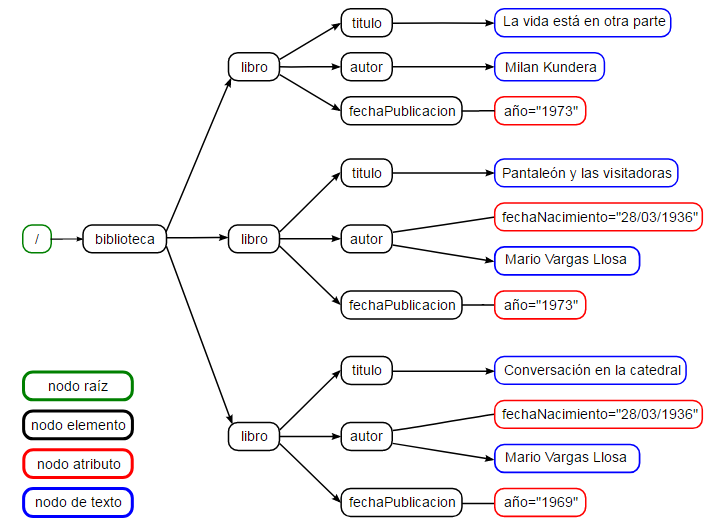
<autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor>

<fechaPublicacion año="1969"/>

</libro>

</biblioteca>

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 elemento y para referirse a los atributos o al texto se utilizan notaciones especiales.

# Sintaxis de las 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, que se explica en esta lección
* 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: selecciona nodos elemento o atributo basándose en sus nombres.
* predicado: restringe la selección del eje a que los nodos cumplan ciertas condiciones.
* selección de nodos: de los nodos seleccionados por el eje y predicado, selecciona los elementos, el texto que contienen o ambos.

# 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"/>**

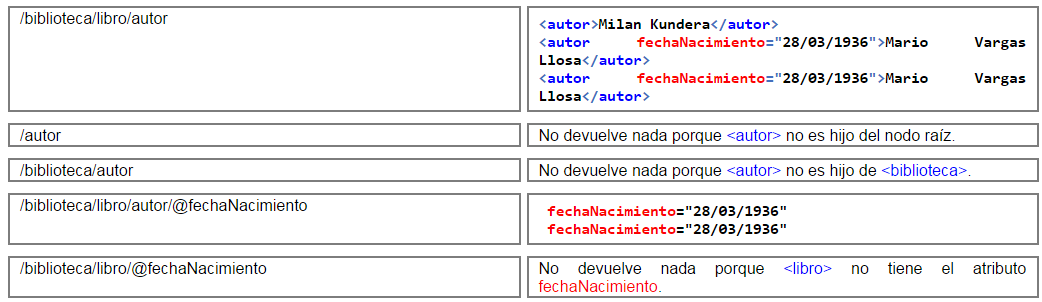
**</libro>**

**</biblioteca>**

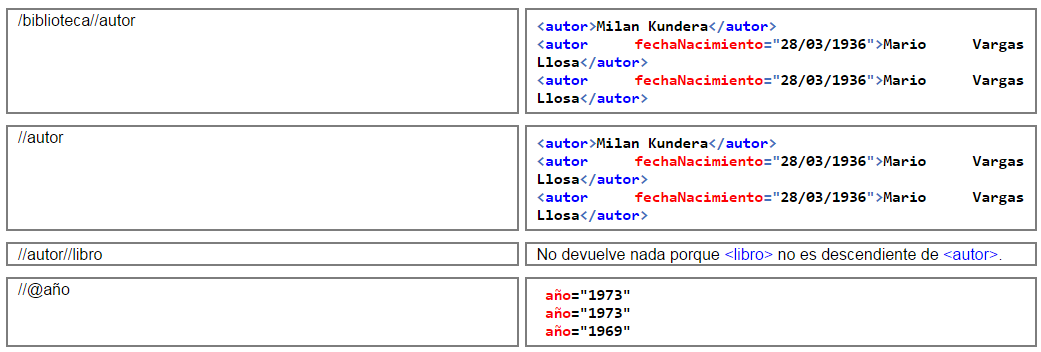
## Eje (en inglés, axis)

El eje nos permite seleccionar un subconjunto de nodos del documento y corresponde a recorridos en el árbol del documento. Los nodos elemento se indican mediante el nombre del elemento. Los nodos atributo se indican mediante @ y el nombre del atributo.

* **/**: si está al principio de la expresión, indica el nodo raíz, si no, indica "hijo". Debe ir seguida del nombre de un elemento

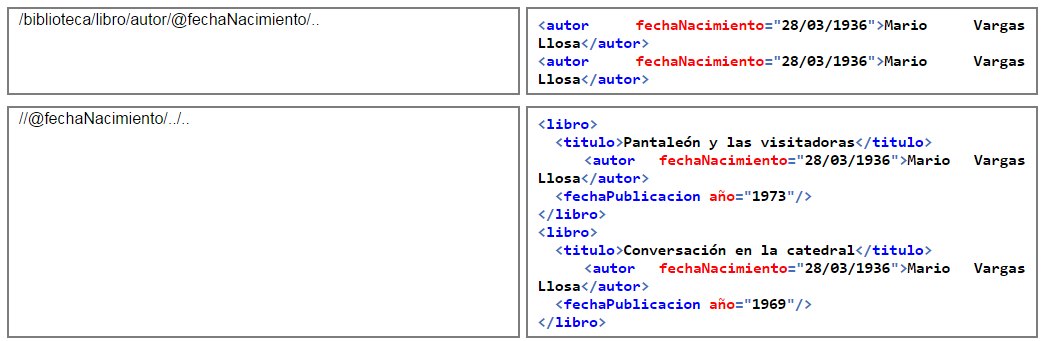


* **//**: indica "descendiente" (hijos, hijos de hijos, etc.)

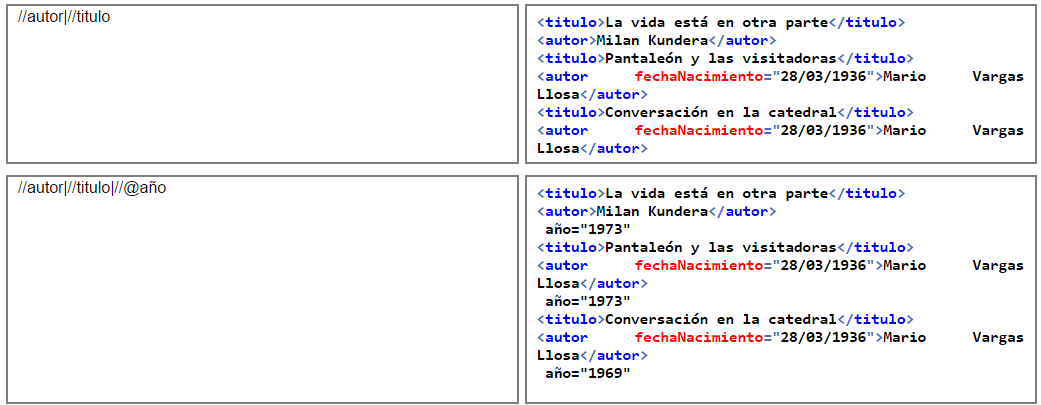


* **/..**: indica el elemento padre

Nota: En el resultado de los ejemplos siguientes se obtienen únicamente los nodos que tienen el atributo fechaNacimiento



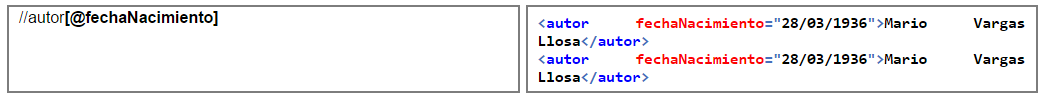
* **|**: permite indicar varios recorridos.



## Predicado (en inglés, predicate)

El predicado se escribe entre corchetes, a continuación del eje. Si el eje ha seleccionado unos nodos, el predicado permite restringir esa selección a los que cumplan determinadas condiciones

* **[@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



* **[condicion]**: selecciona los nodos que cumplen la condición.

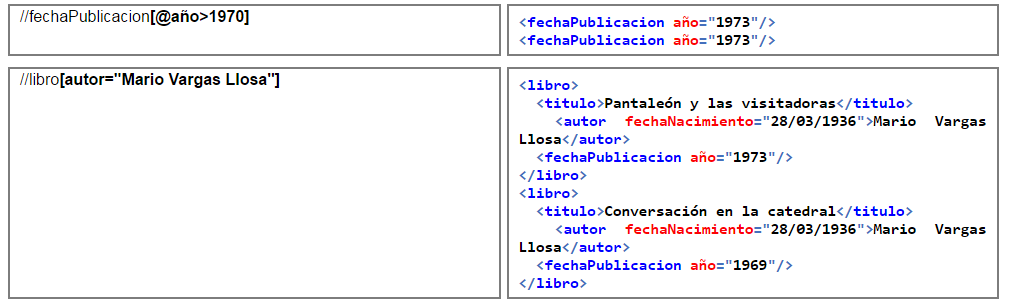
Los predicados permiten definir condiciones sobre los valores de los atributos. En las condiciones se pueden utilizar los operadores siguientes:

* operadores lógicos: and, or, not()
* operadores aritméticos: +, -, \*, div, mod
* operadores de comparación: =, !=, <, >, <=, >=

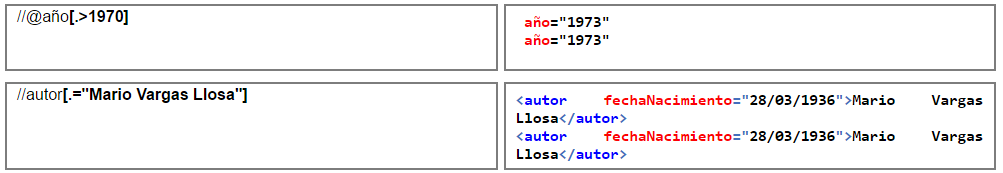
Las comparaciones se pueden hacer entre valores de nodos y atributos o con cadenas de texto o numéricas. En el caso de las cadenas de texto deben estar rodeadeas por comillas simples o dobles. En el caso de las cadenas numéricas, las comillas son optativas.

* La condición puede utilizar el valor de un atributo (utilizando @) o el texto que contiene el elemento.

En los ejemplos siguientes se obtienen respectivamente los elementos <fechaPublicacion> cuyo atributo año es posterior/mayor a 1970 y los elementos <libro> cuyo subelemento <autor> tiene como contenido "Mario Vargas LLosa":

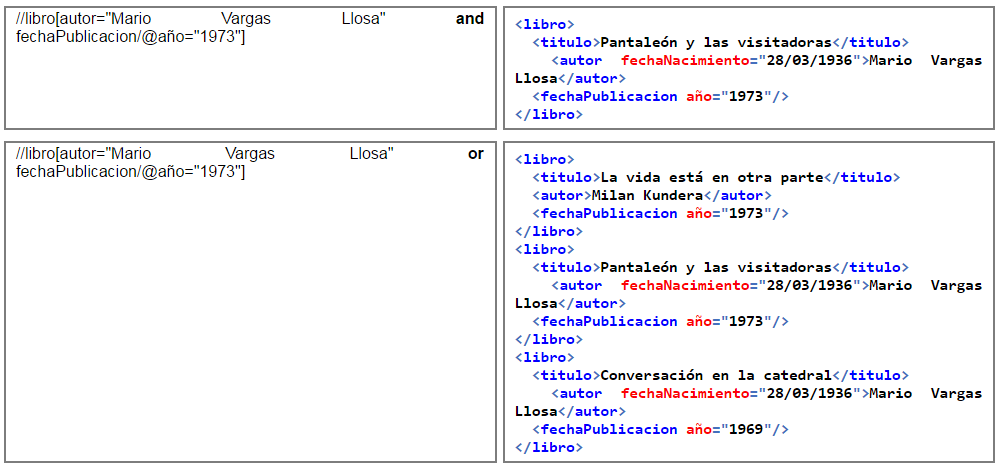


Para hacer referencia al propio valor del elemento seleccionado se utiliza el punto (.).



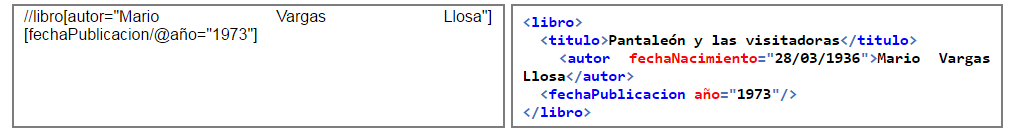
Un predicado puede contener condiciones compuestas.

En los ejemplos siguientes se seleccionan, respectivamente, los libros escritos por Mario Vargas Llosa y publicados en 1973 (primer ejemplo) y los libros escritos por Mario Vargas Llosa o publicados en 1973 (segundo ejemplo):



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

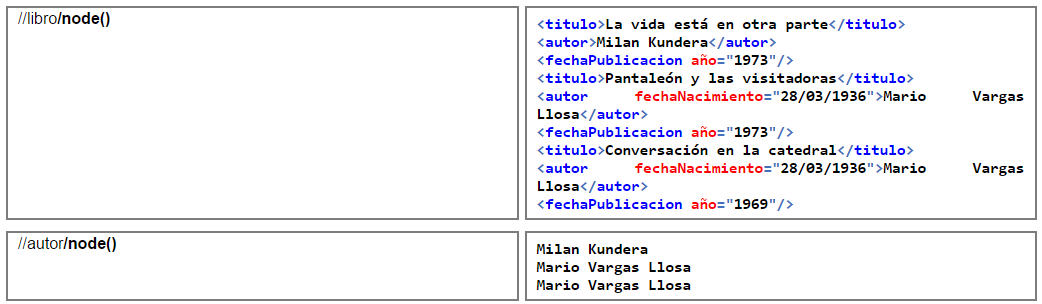
En el ejemplo siguiente se seleccionan los libros escritos por Mario Vargas Llosa y publicados en 1973:

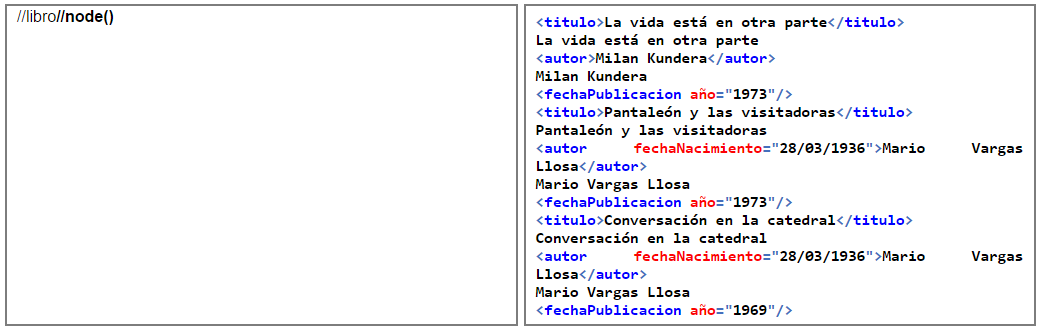


## Selección de nodos (en inglés, node test)

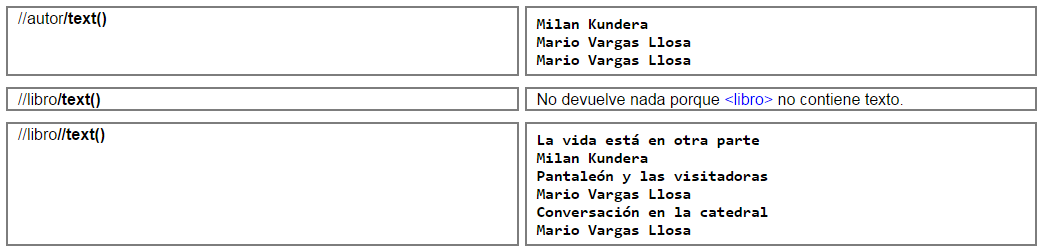
La selección de nodos se escribe a continuación del eje y el predicado. Si el eje y el predicado han seleccionado unos nodos, la selección de nodos indica con qué parte de esos nodos nos quedamos.

* **/node():** selecciona todos los hijos (elementos o texto) del nodo.
* **//node():** selecciona todos los descendientes (elementos o texto) del nodo.

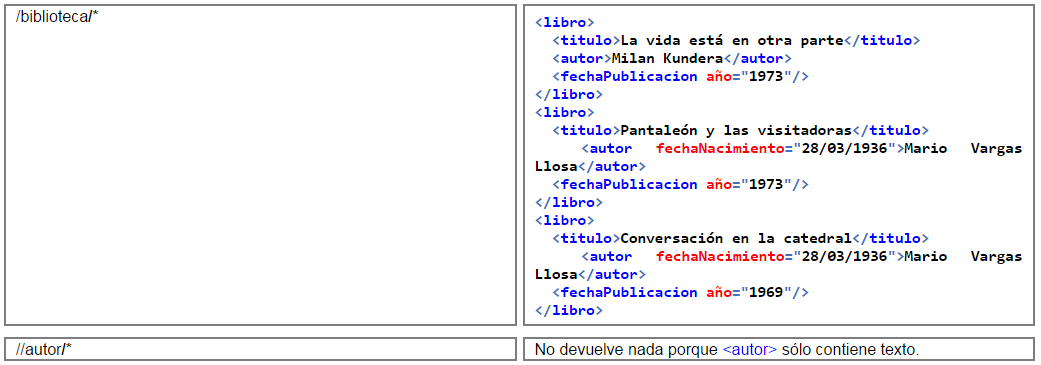




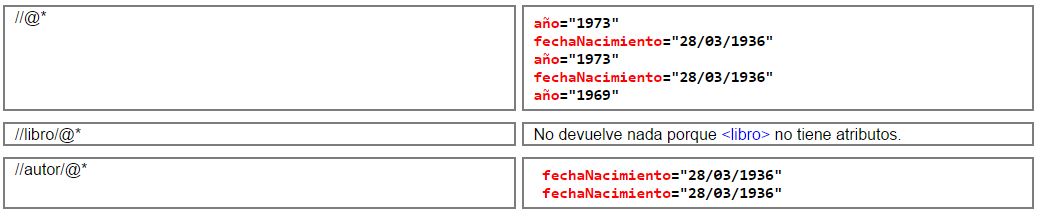
* **/text():** selecciona únicamente el texto contenido en el nodo.
* **//text():** selecciona únicamente el texto contenido en el nodo y todos sus descendientes.



* **/\***: selecciona todos los hijos (sólo elementos) del nodo.
* **//\***: selecciona todos los descendientes (sólo elementos) del nodo.



* **/@\*: selecciona todos los atributos del nodo.**
* **//@\*: selecciona todos los atributos de los descendientes del nodo.**



# Pasos de búsqueda consecutivos

Una expresión XPath puede contener varios pasos de búsqueda consecutivos. Cada uno incluirá su eje (y en su caso, su predicado) y el último paso de búsqueda incluirá en su caso una selección de nodos. Cada paso de búsqueda trabaja a partir de los nodos seleccionados por el paso de búsqueda anterior.

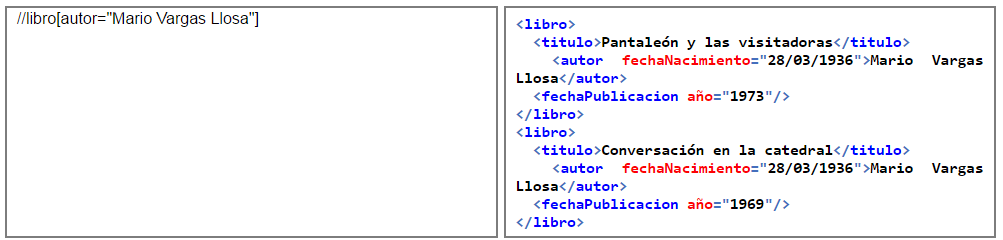
En el ejemplo siguiente se obtienen los títulos de los libros publicados después de 1970, mediante dos pasos de búsqueda:

* en el primer paso (//fechaPublicacion[@año>1970]) se seleccionan los elementos <fechaPublicacion> cuyo atributo año es superior a 1970.
* en el segundo paso (/../titulo), se seleccionan primero los elementos padre (/..) de los <fechaPublicacion> seleccionados en el primer paso de búsqueda (es decir, elementos <libro>) y a continuación sus subelementos <titulo>.

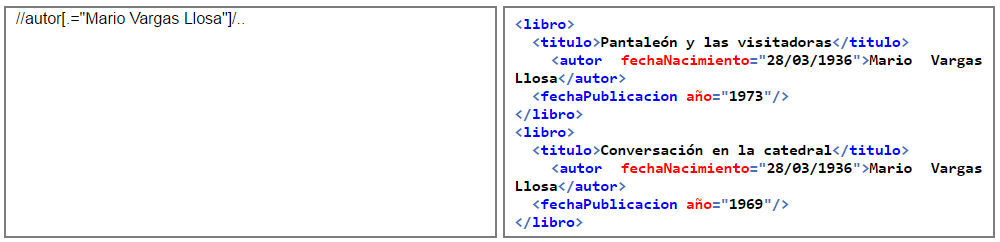


Un determinado resultado se puede obtener mediante un sólo paso de búsqueda o mediante varios pasos.

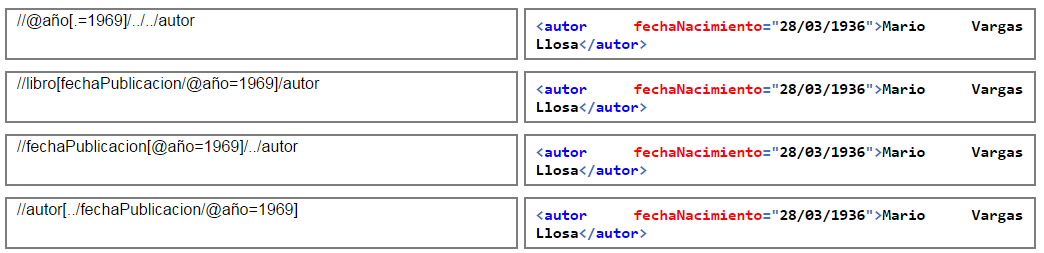
* En los ejemplos siguientes se obtienen los libros escritos por Mario Vargas Llosa de dos formas distintas:
  + mediante un sólo paso de búsqueda. Se seleccionan los elementos <libro> cuyo subelemento <autor> tiene como contenido la cadena "Mario Vargas Llosa".



* + mediante dos pasos de búsqueda. En el primer paso se seleccionan los elementos <autor> cuyo contenido es la cadena "Mario Vargas Llosa". En el segundo paso de búsqueda se seleccionan los elementos padre (es decir, los elementos <libro>).



En los ejemplos siguientes se obtiene el autor que ha publicado libros en 1969 de varias formas distintas:



# 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":



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



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



Como cada una de las expresiones puede escribirse de varias maneras, en realidad hay muchas formas de encontrar la respuesta. Por ejemplo, en la solución siguiente los predicados se encuentran al final del eje en cada subexpresión:



# Ejercicios

Dado el siguiente documento XML, escriba las expresiones XPath que devuelvan la respuesta deseada.

**<?xml version="1.0" encoding="UTF-8"?>**

**<ies>**

**<nombre>IES Abastos</nombre>**

**<web>http://www.iesabastos.org</web>**

**<ciclos>**

**<ciclo id="ASIR">**

**<nombre>Administración de Sistemas Informáticos en Red</nombre>**

**<grado>Superior</grado>**

**<decretoTitulo año="2009" />**

**</ciclo>**

**<ciclo id="DAW">**

**<nombre>Desarrollo de Aplicaciones Web</nombre>**

**<grado>Superior</grado>**

**<decretoTitulo año="2010" />**

**</ciclo>**

**<ciclo id="SMR">**

**<nombre>Sistemas Microinformáticos y Redes</nombre>**

**<grado>Medio</grado>**

**<decretoTitulo año="2008" />**

**</ciclo>**

**</ciclos>**

**</ies>**

1. Nombre del Instituto: <nombre>IES Abastos</nombre>
2. Página web del Instituto: <http://www.iesabastos.org>
3. Nombre de los Ciclos Formativos:

Administración de Sistemas Informáticos en Red

Desarrollo de Aplicaciones Web

Sistemas Microinformáticos y Redes

1. Siglas por las que se conocen los Ciclos Formativos:

id="ASIR"

id="DAW"

id="SMR

1. Años en los que se publicaron los decretos de título de los Ciclos Formativos:

año="2009"

año="2010"

año="2008"

1. Ciclos Formativos de Grado Medio (se trata de obtener el elemento <ciclo> completo):

Nota: Resuelva este ejercicio de dos formas distintas, en un único paso de búsqueda y en dos pasos de búsqueda.

<ciclo id="SMR">

<nombre>Sistemas Microinformáticos y Redes</nombre>

<grado>Medio</grado>

<decretoTitulo año="2008"/>

</ciclo>

1. Nombre de los Ciclos Formativos anteriores a 2010:

Nota: Resuelva este ejercicio de dos formas distintas, en un único paso de búsqueda y en dos pasos de búsqueda.

Administración de Sistemas Informáticos en Red

Sistemas Microinformáticos y Redes

1. Nombre de los Ciclos Formativos de 2008 o 2010:

Nota: Resuelva este ejercicio de dos formas distintas, en un único paso de búsqueda y en dos pasos de búsqueda.

Desarrollo de Aplicaciones Web

Sistemas Microinformáticos y Redes

**Dado el siguiente XML**

**<?xml version="1.0" encoding="UTF-8"?>**

**<ies>**

**<modulos>**

**<modulo id="0228">**

**<nombre>Aplicaciones web</nombre>**

**<curso>2</curso>**

**<horasSemanales>4</horasSemanales>**

**<ciclo>SMR</ciclo>**

**</modulo>**

**<modulo id="0372">**

**<nombre>Gestión de bases de datos</nombre>**

**<curso>1</curso>**

**<horasSemanales>5</horasSemanales>**

**<ciclo>ASIR</ciclo>**

**</modulo>**

**<modulo id="0373">**

**<nombre>Lenguajes de marcas y sistemas de gestión de información</nombre>**

**<curso>1</curso>**

**<horasSemanales>3</horasSemanales>**

**<ciclo>ASIR</ciclo>**

**<ciclo>DAW</ciclo>**

**</modulo>**

**<modulo id="0376">**

**<nombre>Implantación de aplicaciones web</nombre>**

**<curso>2</curso>**

**<horasSemanales>5</horasSemanales>**

**<ciclo>ASIR</ciclo>**

**</modulo>**

**</modulos>**

**</ies>**

1. Nombre de los módulos del ciclo ASIR:

Gestión de bases de datos

Lenguajes de marcas y sistemas de gestión de información

Implantación de aplicaciones web

1. Nombre de los módulos que se imparten en el segundo curso de cualquier ciclo:

Aplicaciones web

Implantación de aplicaciones web

1. Nombre de los módulos que se imparten en el primer curso de ASIR:

Gestión de bases de datos

Lenguajes de marcas y sistemas de gestión de información

1. Horas semanales de los módulos de más de 3 horas semanales:

4

5

5

**Dado este XML:**

**<?xml version="1.0" encoding="UTF-8"?>**

**<ies>**

**<nombre>IES Abastos</nombre>**

**<web>http://www.iesabastos.org</web>**

**<ciclos>**

**<ciclo id="ASIR">**

**<nombre>Administración de Sistemas Informáticos en Red</nombre>**

**<grado>Superior</grado>**

**<decretoTitulo año="2009" />**

**</ciclo>**

**<ciclo id="DAW">**

**<nombre>Desarrollo de Aplicaciones Web</nombre>**

**<grado>Superior</grado>**

**<decretoTitulo año="2010" />**

**</ciclo>**

**<ciclo id="SMR">**

**<nombre>Sistemas Microinformáticos y Redes</nombre>**

**<grado>Medio</grado>**

**<decretoTitulo año="2008" />**

**</ciclo>**

**</ciclos>**

**<modulos>**

**<modulo id="0228">**

**<nombre>Aplicaciones web</nombre>**

**<curso>2</curso>**

**<horasSemanales>4</horasSemanales>**

**<ciclo>SMR</ciclo>**

**</modulo>**

**<modulo id="0372">**

**<nombre>Gestión de bases de datos</nombre>**

**<curso>1</curso>**

**<horasSemanales>5</horasSemanales>**

**<ciclo>ASIR</ciclo>**

**</modulo>**

**<modulo id="0373">**

**<nombre>Lenguajes de marcas y sistemas de gestión de información</nombre>**

**<curso>1</curso>**

**<horasSemanales>3</horasSemanales>**

**<ciclo>ASIR</ciclo>**

**<ciclo>DAW</ciclo>**

**</modulo>**

**<modulo id="0376">**

**<nombre>Implantación de aplicaciones web</nombre>**

**<curso>2</curso>**

**<horasSemanales>5</horasSemanales>**

**<ciclo>ASIR</ciclo>**

**</modulo>**

**</modulos>**

**</ies>**

1. Nombre de los módulos del ciclo "Sistemas Microinformáticos y Redes" (en la expresión final no deben aparecer las siglas SMR):

Aplicaciones web

1. Nombre de los ciclos que incluyen el módulo "Lenguajes de marcas y sistemas de gestión de información":

Administración de Sistemas Informáticos en Red

Desarrollo de Aplicaciones Web

1. Nombre de los módulos de ciclos de Grado Superior:

Gestión de bases de datos

Lenguajes de marcas y sistemas de gestión de información

Implantación de aplicaciones web

1. Nombre de los módulos de ciclos cuyo título se aprobó en 2008:

Aplicaciones web

1. Grado de los ciclos con módulos de primer curso:

Superior

Superior

Software, Herramienta online:

http://videlibri.sourceforge.net/cgi-bin/xidelcgi

# EJERCICIOS DE REPASO CON SOLUCIONES:

[**http://www.mclibre.org/consultar/xml/ejercicios/xpath.html**](http://www.mclibre.org/consultar/xml/ejercicios/xpath.html)

[**https://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&cad=rja&uact=8&ved=0ahUKEwjpqe707PfTAhWPOsAKHTE9BYcQFgg3MAM&url=http%3A%2F%2Fdis.um.es%2F~lopezquesada%2Fdocumentos%2FIES\_1213%2FLMSGI%2Fcurso%2FUT8%2Fejerciciosxpath.doc&usg=AFQjCNETDD6Fao6eoM8eOrHKEisk\_KnQxA&sig2=nXdZDO347iHgzyMYb7RJeA**](https://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&cad=rja&uact=8&ved=0ahUKEwjpqe707PfTAhWPOsAKHTE9BYcQFgg3MAM&url=http%3A%2F%2Fdis.um.es%2F~lopezquesada%2Fdocumentos%2FIES_1213%2FLMSGI%2Fcurso%2FUT8%2Fejerciciosxpath.doc&usg=AFQjCNETDD6Fao6eoM8eOrHKEisk_KnQxA&sig2=nXdZDO347iHgzyMYb7RJeA)

**Fuente:**

[**http://www.mclibre.org**](http://www.mclibre.org)