

Ejercicios propuestos

Tema 10. Gestión y recuperación de información. XPath

Ejercicio 10.1

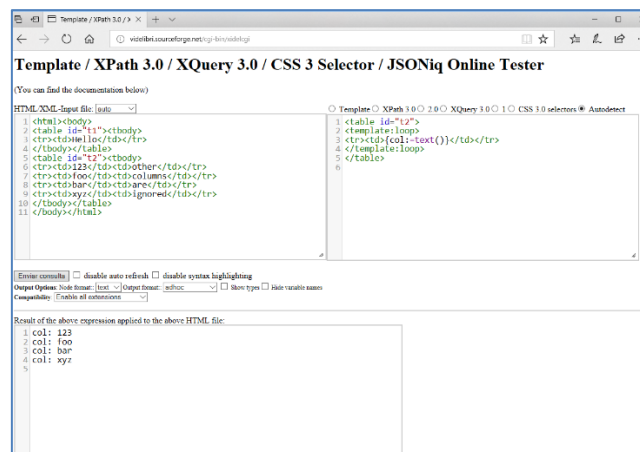
Uso de herramientas on-line para realizar consultas XPath

Ejercicio guiado. En este ejercicio se utiliza la **herramienta on-line videlibri** para practicar las búsquedas XPath en un documento XML. Aunque en ejercicios posteriores se utilizan diversas herramientas de escritorio para probar expresiones de consulta XPath, se propone también utilizar la herramienta on-line videlibri, o cualquier otra similar, por su facilidad de uso.

Crea el siguiente documento XML utilizando XML Copy Editor o cualquier otro editor de código XML.

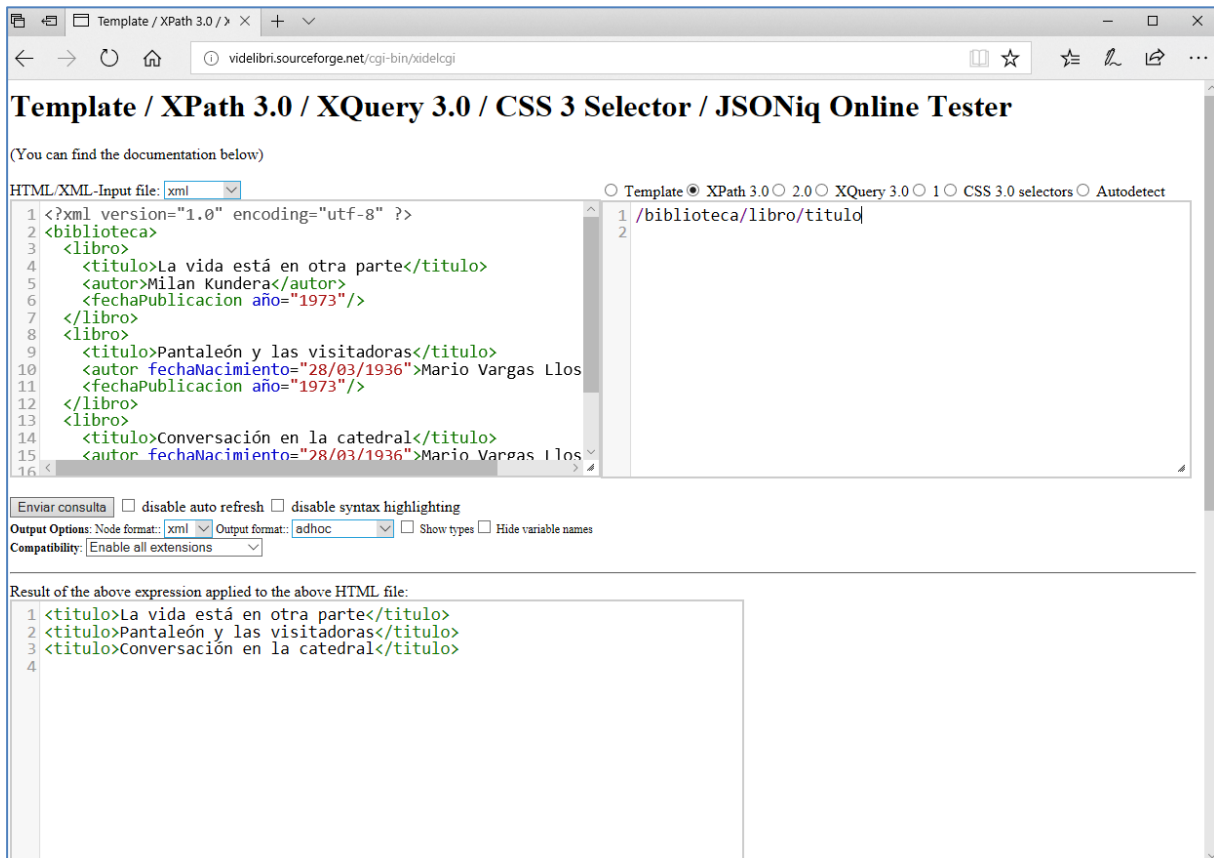
```
<?xml version="1.0" encoding="utf-8" ?>
<biblioteca>
  <libro>
    <titulo>La vida está en otra parte</titulo>
    <autor>Milan Kundera</autor>
    <fechaPublicacion año="1973"/>
  </libro>
  <libro>
    <titulo>Pantaleón y las visitadoras</titulo>
    <autor fechaNacimiento="28/03/1936">Mario Vargas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

A continuación, visita la página: <http://videlibri.sourceforge.net/cgi-bin/xidelcgi>.



En la interfaz web de la página anterior pueden apreciarse tres zonas bien diferenciadas: la zona de documento XML, la zona de expresión de consulta XPath y la zona de resultados. Realiza las siguientes acciones para procesar una expresión de consulta XPath sobre un documento XML.

1. En el área de texto situada en la parte superior izquierda:
 - a. Elimina el contenido incluido por defecto y pega el contenido del documento XML de ejemplo creado anteriormente sobre el cual se va a realizar una consulta XPath.
 - b. En el cuadro combinado **HTML/XML-Input file**, selecciona la opción **xml**.
2. Selecciona la opción **xml** en el cuadro combinado **Node format** que está situado en el grupo de características **Output Options** debajo del área de texto anterior.
3. En el área de texto situada en la parte superior derecha:
 - a. Marca la casilla de verificación **XPath 3.0**.
 - b. Pega o introduce la expresión XPath que se va a procesar: **/biblioteca/libro/titulo**
4. En el área de texto situada en la parte inferior izquierda se mostrará el resultado de la expresión de consulta XPath introducida sobre el documento XML del ejemplo.



Template / XPath 3.0 / XQuery 3.0 / CSS 3 Selector / JSONiq Online Tester

(You can find the documentation below)

HTML/XML-Input file:

☐ Template ☒ XPath 3.0 ☐ 2.0 ☐ XQuery 3.0 ☐ 1 ☐ CSS 3.0 selectors ☐ Autodetect

```
1 <?xml version="1.0" encoding="utf-8" ?>
2 <biblioteca>
3   <libro>
4     <titulo>La vida está en otra parte</titulo>
5     <autor>Milan Kundera</autor>
6     <fechaPublicacion año="1973"/>
7   </libro>
8   <libro>
9     <titulo>Pantaleón y las visitadoras</titulo>
10    <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor>
11    <fechaPublicacion año="1973"/>
12  </libro>
13  <libro>
14    <titulo>Conversación en la catedral</titulo>
15    <autor fechaNacimiento="28/03/1936">Mario Vargas Llosa</autor>
16  </libro>
```

Enviar consulta ☐ disable auto refresh ☐ disable syntax highlighting

Output Options: Node format: Output format: ☐ Show types ☐ Hide variable names

Compatibility:

Result of the above expression applied to the above HTML file:

```
1 <titulo>La vida está en otra parte</titulo>
2 <titulo>Pantaleón y las visitadoras</titulo>
3 <titulo>Conversación en la catedral</titulo>
4
```

Pueden también utilizarse otras herramientas on-line como:

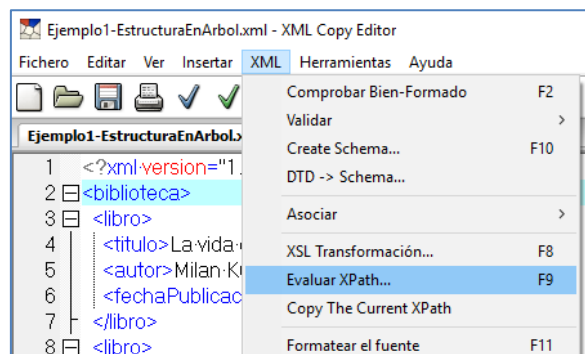
- XPath Tester: <http://www.xpathtester.com/xpath>
- FreeFormatter: <https://www.freeformatter.com/xpath-tester.html>
- Code Beauty: <https://codebeautify.org/Xpath-Tester>

Ejercicio 10.2

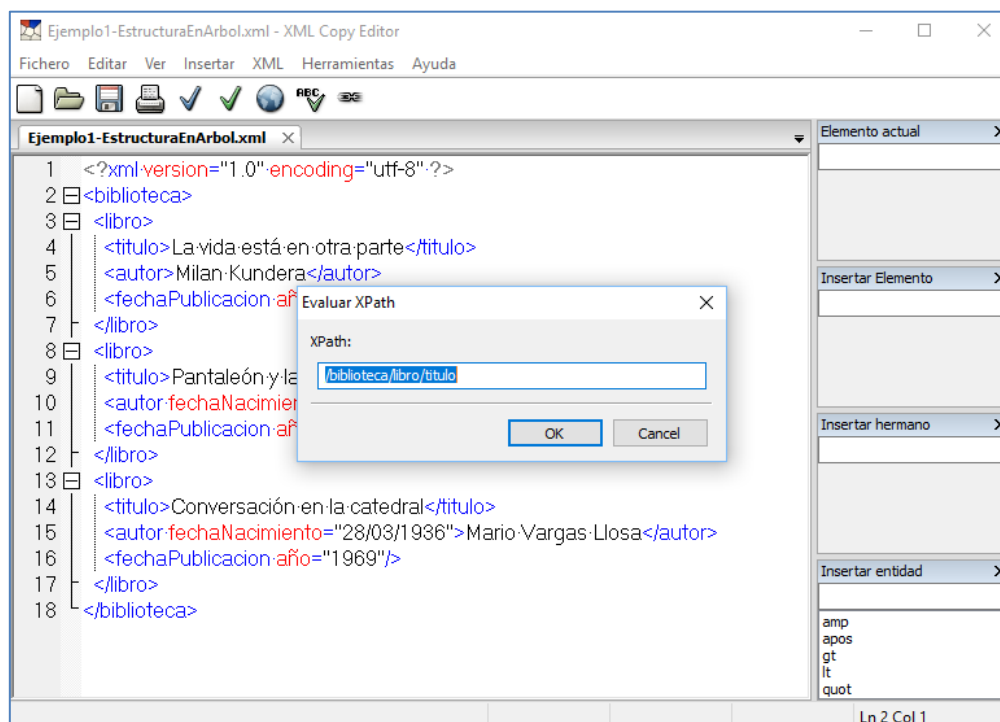
Uso de XML Copy Editor para realizar consultas XPath

Ejercicio guiado. En este ejercicio se utiliza el editor de código **XML Copy Editor** para practicar las búsquedas XPath en un documento XML. Aunque, preferiblemente, se utilizará la herramienta avanzada BaseX, el editor XML Copy Editor incorpora un procesador básico de XPath para evaluar consultas aisladas de XPath de una forma sencilla. El resultado se muestra en otra ventana del editor.

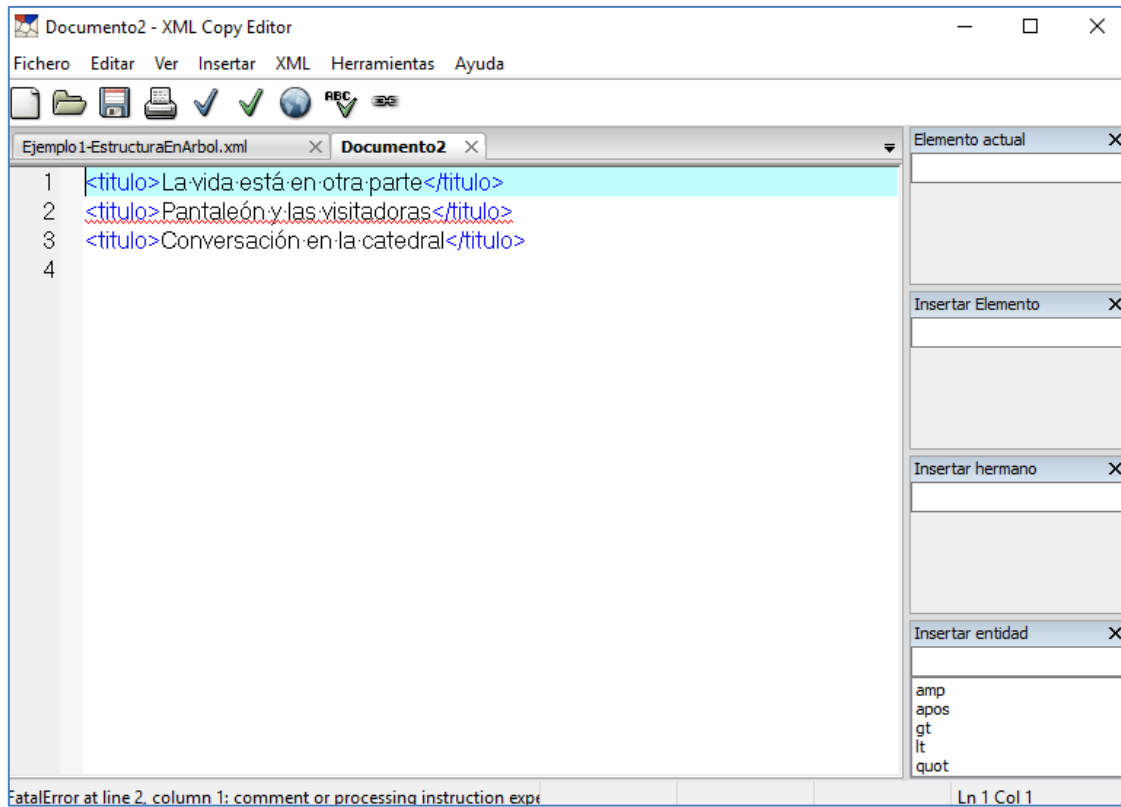
Inicia XML Copy Editor y abre el documento XML de ejemplo creado en el ejercicio anterior. A continuación, selecciona la opción **Evaluar XPath...** del menú **XML**. Esta opción permite introducir la expresión de consulta XPath a aplicar sobre el documento XML actual.



En el cuadro de diálogo **Evaluar XPath** introduce la expresión de consulta, que en este caso es: `/biblioteca/libro/titulo`. Y, a continuación, haz clic en el botón **Ok**.



En otra ventana del editor se mostrará el resultado de la expresión de consulta XPath introducida.

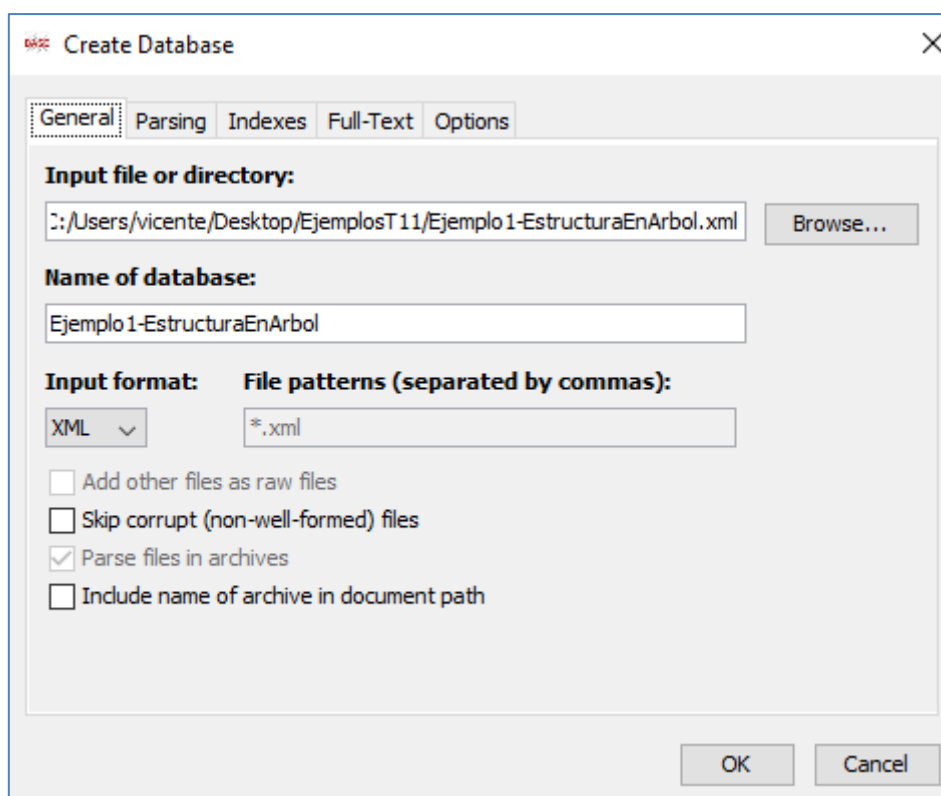


Ejercicio 10.3

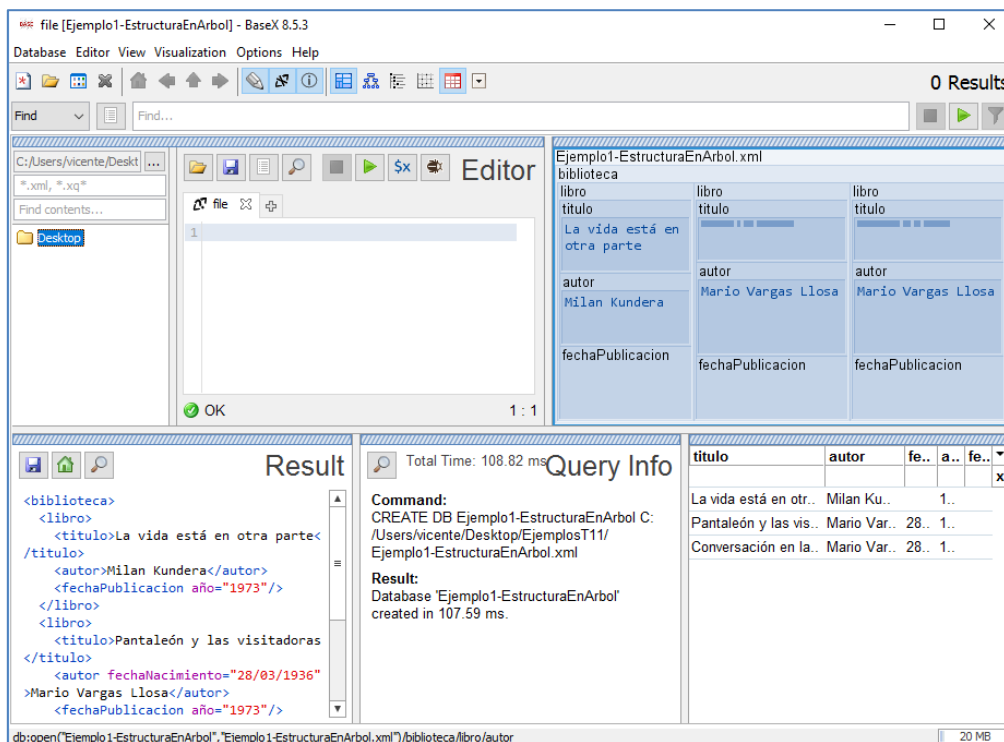
Uso de BaseX para realizar consultas XPath

Ejercicio guiado. En este ejercicio se utiliza el sistema gestor de base de datos XML nativa **BaseX** para practicar las búsquedas XPath en un documento XML. Se utiliza esta herramienta avanzada para evaluar un ejemplo de consulta de XPath sobre el mismo documento XML del ejercicio anterior.

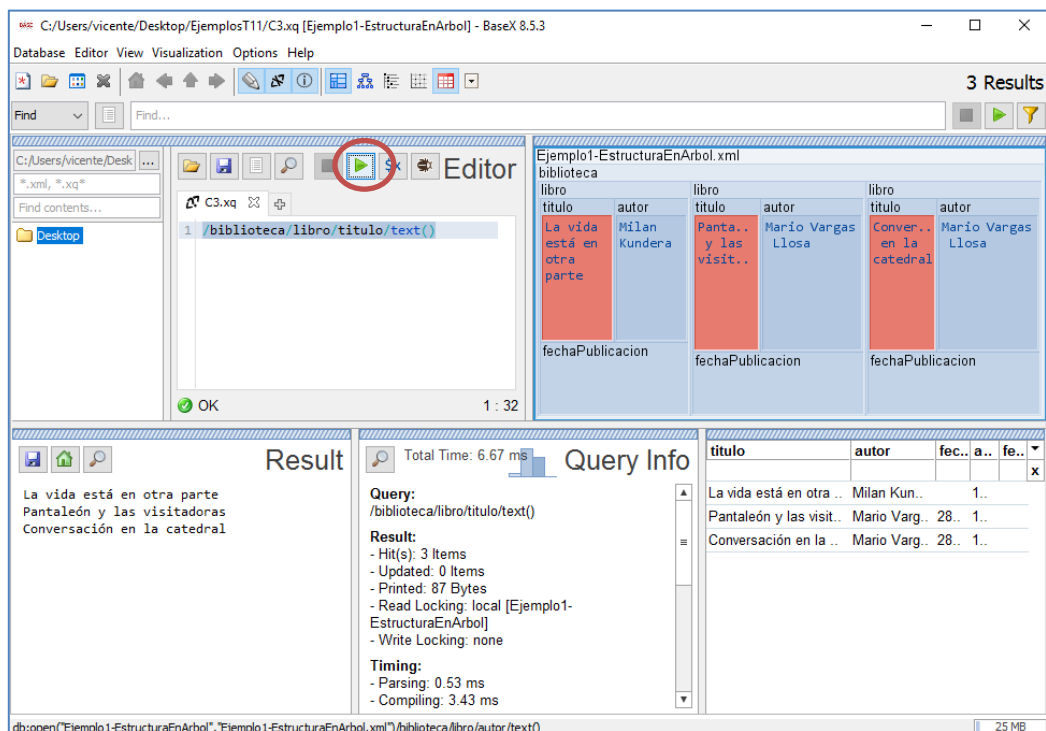
Una vez iniciado BaseX, en primer lugar, se crea una base de datos XML en base al documento XML que va a utilizarse para hacer la consulta. Para ello, seleccionar la opción **New...** del menú **Database** y, a continuación, en la ficha General del cuadro de diálogo **Create Database** seleccionar el archivo correspondiente haciendo clic en el botón **Browse**. Es importante comprobar que el valor de la opción **Input format** sea XML, para indicar que se utiliza un documento XML como origen de datos.



Una vez realizada la acción anterior, se habrá creado la base de datos XML. En la siguiente ilustración puede apreciarse la ventana de ejecución de BaseX una vez que se ha creado la base de datos. Puede apreciarse que incluye, a su vez, diversas ventanas. La ventana **Editor** permite introducir las expresiones de consulta a evaluar y ejecutar contra la base de datos creada. La ventana de su derecha muestra una representación del documento XML en forma de Mapa. La ventana **Result** muestra el resultado de la última consulta realizada en formato de texto. En este caso, se muestra la estructura jerárquica del documento XML al completo en formato de texto, porque aún no se ha ejecutado ninguna consulta. La ventana **Query Info** muestra información sobre la consulta realizada. La ventana de su derecha muestra una representación del documento XML en forma de Tabla.



Para realizar una consulta, introducir la expresión XPath correspondiente en la ventana Editor y hacer clic en el botón **Run Query** que se representa mediante una flecha verde. Al introducir la expresión: `/biblioteca/libro/titulo/text()` y ejecutar la consulta, se obtiene el siguiente resultado.



Ejercicio 10.4

Uso de expresiones XPath

Dado el siguiente documento XML, denominado *Ejercicio10-4.xml*, escribe las expresiones XPath correspondientes que permitan devolver los resultados deseados que se muestran los cuadros.

```
<?xml version="1.0" encoding="utf-8" ?>
<centroFormacion>
  <nombre localidad="Alicante">IES Mare Nostrum</nombre>
  <web>http://www.iesmarenostrum.com</web>
  <correo>correo@iesmarenostrum.com</correo>
  <ciclos>
    <ciclo id="ASIR" familia="Informática y Comunicaciones">
      <nombre>Administración de Sistemas Informáticos en Red</nombre>
      <grado>Superior</grado>
      <decretoTitulo año="2009" />
    </ciclo>
    <ciclo id="DAW" familia="Informática y Comunicaciones">
      <nombre>Desarrollo de Aplicaciones Web</nombre>
      <grado>Superior</grado>
      <decretoTitulo año="2010" />
    </ciclo>
    <ciclo id="SMR" familia="Informática y Comunicaciones">
      <nombre>Sistemas Microinformáticos y Redes</nombre>
      <grado>Medio</grado>
      <decretoTitulo año="2008" />
    </ciclo>
    <ciclo id="IO" familia="Informática y Comunicaciones">
      <nombre>Informática de Oficina</nombre>
      <grado>FP básica</grado>
      <decretoTitulo año="2014" />
    </ciclo>
    <ciclo id="AF" familia="Administración y Gestión">
      <nombre>Administración y Finanzas</nombre>
      <grado>Superior</grado>
      <decretoTitulo año="2011" />
    </ciclo>
    <ciclo id="AD" familia="Administración y Gestión">
      <nombre>Asistencia a la Dirección</nombre>
      <grado>Superior</grado>
      <decretoTitulo año="2011" />
    </ciclo>
    <ciclo id="GA" familia="Administración y Gestión">
      <nombre>Gestión Administrativa</nombre>
      <grado>Medio</grado>
      <decretoTitulo año="2009" />
    </ciclo>
  </ciclos>
</centroFormacion>
```

- Obtener el nombre del centro de formación.

Resultado deseado	<code><nombre localidad="Alicante">IES Mare Nostrum</nombre></code>
--------------------------	---

- Obtener el valor de la página web del centro de formación.

Resultado deseado	<code>http://www.iesmarenostrum.com</code>
--------------------------	--

- Obtener el valor del nombre de los ciclos formativos impartidos.

Resultado deseado	Administración de Sistemas Informáticos en Red Desarrollo de Aplicaciones Web Sistemas Microinformáticos y Redes Informática de Oficina Administración y Finanzas Asistencia a la Dirección Gestión Administrativa
--------------------------	--

- Obtener el valor de las siglas de los ciclos formativos impartidos.

Resultado deseado	ASIR DAW SMR IO AF AD GA
--------------------------	--

- Obtener los elementos <ciclo> (completos) de los ciclos formativos de grado medio.

Resultado deseado	<code><ciclo id="SMR" familia="Informática y Comunicaciones"> <nombre>Sistemas Microinformáticos y Redes</nombre> <grado>Medio</grado> <decretoTitulo año="2008"/> </ciclo> <ciclo id="GA" familia="Administración y Gestión"> <nombre>Gestión Administrativa</nombre> <grado>Medio</grado> <decretoTitulo año="2009"/> </ciclo></code>
--------------------------	---

- Obtener los elementos <nombre> de los ciclos formativos de grado superior.

Resultado deseado	<nombre>Administración de Sistemas Informáticos en Red</nombre> <nombre>Desarrollo de Aplicaciones Web</nombre> <nombre>Administración y Finanzas</nombre> <nombre>Asistencia a la Dirección</nombre>
--------------------------	--

- Obtener el valor de los elementos <nombre> de los ciclos formativos anteriores a 2010.

Resultado deseado	Administración de Sistemas Informáticos en Red Sistemas Microinformáticos y Redes Gestión Administrativa
--------------------------	--

- Obtener los elementos <nombre> de los ciclos de 2009 y 2010.

Resultado deseado	<nombre>Administración de Sistemas Informáticos en Red</nombre> <nombre>Desarrollo de Aplicaciones Web</nombre> <nombre>Gestión Administrativa</nombre>
--------------------------	---

- Obtener el valor de los elementos <nombre> de los ciclos formativos de la familia profesional de Informática y comunicaciones.

Resultado deseado	Administración de Sistemas Informáticos en Red Desarrollo de Aplicaciones Web Sistemas Microinformáticos y Redes Informática de Oficina
--------------------------	--

- Obtener los valores de los atributos año en los que se publicaron los títulos de todos los ciclos formativos.

Resultado deseado	2009 2010 2008 2014 2011 2011 2009
--------------------------	--

Ejercicio 10.5

Uso de expresiones XPath

Dado el siguiente documento XML, denominado *Ejercicio10-5.xml*, escribe las expresiones XPath correspondientes que permitan devolver los resultados deseados que se muestran en los cuadros.

```
<?xml version="1.0" encoding="utf-8" ?>
<bookstore specialty="novel">
  <book style="autobiography">
    <author>
      <first-name>Joe</first-name>
      <last-name>Bob</last-name>
      <award>Trenton Literary Review Honorable Mention</award>
    </author>
    <price>12</price>
  </book>
  <book style="textbook">
    <author>
      <first-name>Mary</first-name>
      <last-name>Bob</last-name>
      <publication>
        Selected Short Stories of
        <first-name>Mary</first-name>
        <last-name>Bob</last-name>
      </publication>
    </author>
    <editor>
      <first-name>Britney</first-name>
      <last-name>Bob</last-name>
    </editor>
    <price>55</price>
  </book>
  <magazine style="glossy" frequency="monthly">
    <price>2.50</price>
    <subscription price="24" per="year"/>
  </magazine>
  <book style="novel" id="myfave">
    <author>
      <first-name>Toni</first-name>
      <last-name>Bob</last-name>
      <degree from="Trenton U">B.A.</degree>
      <degree from="Harvard">Ph.D.</degree>
      <award>Pulitzer</award>
      <publication>Still in Trenton</publication>
      <publication>Trenton Forever</publication>
    </author>
    <price intl="Canada" exchange="0.7">6.50</price>
    <excerpt>
      <definition-list>
        <term>Trenton</term>
        <definition>misery</definition>
      </definition-list>
    </excerpt>
  </book>
</bookstore>
```

- Obtener todos los elementos author (completos).

Resultado deseado	<pre> <author> <first-name>Joe</first-name> <last-name>Bob</last-name> <award>Trenton Literary Review Honorable Mention</award> </author> <author> <first-name>Mary</first-name> <last-name>Bob</last-name> <publication>Selected Short Stories of<first-name>Mary</first-name> <last-name>Bob</last-name> </publication> </author> <author> <first-name>Toni</first-name> <last-name>Bob</last-name> <degree from="Trenton U">B.A.</degree> <degree from="Harvard">Ph.D.</degree> <award>Pulitzer</award> <publication>Still in Trenton</publication> <publication>Trenton Forever</publication> </author> </pre>
-------------------	--

- Obtener todos los elementos <book> cuyo valor de atributo style es igual al valor de atributo specialty del elemento <bookstore> en la raíz del documento.

Resultado deseado	<pre> <book style="novel" id="myfave"> <author> <first-name>Toni</first-name> <last-name>Bob</last-name> <degree from="Trenton U">B.A.</degree> <degree from="Harvard">Ph.D.</degree> <award>Pulitzer</award> <publication>Still in Trenton</publication> <publication>Trenton Forever</publication> </author> <price intl="Canada" exchange="0.7">6.50</price> <excerpt> <definition-list> <term>Trenton</term> <definition>misery</definition> </definition-list> </excerpt> </book> </pre>
-------------------	---

- Obtener los elementos <last-name> que son secundarios de los elementos <book>.

Resultado deseado	<pre> <last-name>Bob</last-name> <last-name>Bob</last-name> <last-name>Bob</last-name> <last-name>Bob</last-name> </pre>
--------------------------	--

- Obtener todos los elementos <book> con atributo id.

Resultado deseado	<pre> <book style="novel" id="myfave"> <author> <first-name>Toni</first-name> <last-name>Bob</last-name> <degree from="Trenton U">B.A.</degree> <degree from="Harvard">Ph.D.</degree> <award>Pulitzer</award> <publication>Still in Trenton</publication> <publication>Trenton Forever</publication> </author> <price intl="Canada" exchange="0.7">6.50</price> <excerpt> <definition-list> <term>Trenton</term> <definition>misery</definition> </definition-list> </excerpt> </book> </pre>
--------------------------	---

- Obtener el valor del elemento <first-name> y del elemento <last-name> de todos los elementos <author>.

Resultado deseado	<pre> Joe Bob Mary Bob Toni Bob </pre>
--------------------------	--

- Obtener el valor del atributo style de todos los elementos <book>.

Resultado deseado	<pre> autobiography textbook novel </pre>
--------------------------	---