



Universidad Veracruzana
Facultad de Estadística e Informática



Tema

Documentos XML: Formato y consulta

Elaborado por:

Mtra. Lorena Alonso Ramírez

lonso@uv.mx



Última actualización: mayo 2022

XML

El lenguaje de marcado extensible (XML) es un formato simple basado en texto para representar información estructurada: documentos, datos, configuración, facturas y mucho más.

XML es uno de los formatos más utilizados para compartir información estructurada entre programas, entre personas, entre computadoras y personas, tanto localmente como a través de redes.

Ejemplo de documento xml



- **Etiquetas (tags)** para marcar secciones en los datos: libro, autor, nacionalidad, etc.

- **Elemento**: sección de datos

`<tagname> datos </tagname>`

- Los elementos pueden **anidarse** a varios niveles: elemento/sub-elemento
- Todo documento debe tener un elemento único en el nivel más alto
- Atributos: los elementos pueden tener atributos, se especifican por pares nombre=valor dentro de la etiqueta de inicio de un elemento

Lenguaje de consulta XPath

- XPath es un lenguaje de navegación y selección de partes de documentos
- Está basado en Expresiones de Camino (path expressions):

secuencia de pasos separados por "/"

```
<libreria>
  <libro idAutor="20">
    <categoria>cocina</categoria>
    <titulo lang="en">Everyday Italian</titulo>
    <anio>2005</anio>
    <precio>30.00</precio>
  </libro>
  <libro idAutor="30">
    <categoria>infantil</categoria>
    <titulo lang="en">Harry Potter</titulo>
    <anio>2005</anio>
    <precio>29.99</precio>
  </libro>
  <autor id="20">
    <nombre>Laura De Laurentiis</nombre>
    <nacionalidad>Italiana</nacionalidad>
  </autor>
  <autor id="30">
    <nombre>J. K. Rowling</nombre>
    <nacionalidad>inglesa</nacionalidad>
  </autor>
</libreria>
```

Ejemplo	Resultado
<ul style="list-style-type: none"> • Secuencia de etiquetas: mostrar los títulos de los libros. libreria/libro/titulo 	<pre><titulo lang="en">Everyday Italian</titulo> <titulo lang="en">Harry Potter</titulo></pre>
<ul style="list-style-type: none"> • Predicados de selección: mostrar los autores de nacionalidad Italiana libreria/autor[nacionalidad="Italiana"] 	<pre><autor id="20"> <nombre>Laura De Laurentiis</nombre> <nacionalidad>Italiana</nacionalidad> </autor></pre>
<ul style="list-style-type: none"> • Acceso a atributos, uso de @: título de libros escritos en inglés (lang="en") libreria/libro/titulo[@lang = "en"] 	<pre><titulo lang="en">Everyday Italian</titulo> <titulo lang="en">Harry Potter</titulo></pre>
<ul style="list-style-type: none"> • Acceso a atributos, uso de @: título de libros escritos en inglés (lang="en") libreria/libro/titulo[@lang = "en"] 	<pre><titulo lang="en">Everyday Italian</titulo> <titulo lang="en">Harry Potter</titulo></pre>

Lenguaje de consulta XQuery

- Lenguaje de consulta de datos XML de propósito general
- Basado en expresiones FLWR
 - **For** - from de SQL
 - **Let** - definición de variable (opcional)
 - **Where** - where de SQL
 - **Result** - select SQL

Consulta sin cláusula where

1. Mostrar los títulos de todos los libros
for \$x in libreria/libro
return \$x/titulo

Consulta con filtro (condición where)

2. El título de los libros del año 2010 o antes
for \$x in libreria/libro
where \$x/anio<2010
return \$x/titulo

Mostrar dos etiquetas en el resultado

3. El título y el precio de los libros de la categoría de 'cocina'
for \$x in libreria/libro
where \$x/categoria = "cocina"
return <resp>{\$x/titulo}{\$x/precio}</resp>

Ordenar alfabéticamente el resultado

4. Mostrar el nombre de los autores ordenados alfabéticamente
for \$x in libreria/autor
order by \$x/nombre descending
return \$x/nombre

Uso de referencias entre las etiquetas, combinar información entre etiquetas

5. Mostrar el título de todos los libros y el nombre de su autor
for \$x in libreria/autor, \$y in libreria/libro
where \$x/@id = \$y/@idAutor
return <resp>{\$y/titulo}{\$x/nombre}</resp>

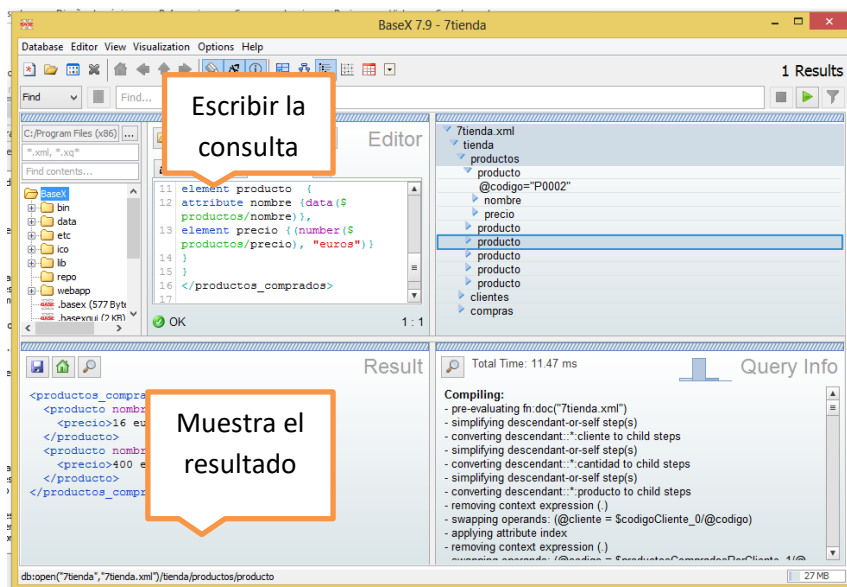
Uso de función count, permite contar cuantas veces aparece una etiqueta

6. Mostrar el nombre de los autores que tengan más de 1 nacionalidad
for \$x in libreria/autor
where count(\$x/nacionalidad) >1
return \$x/nombre

Práctica

Consultar una base de datos de documentos XML

1. Descargar e instalar el programa BaseX
<https://basex.org/download/>
2. Descargar la base de datos 'Mondial' en formato xml
<https://www.dbis.informatik.uni-goettingen.de/Mondial/#XML>
3. Crear una nueva base de datos: Menú database, opción new, seleccionar el archivo xml que se descargó (mondial.xml)



Hacer una exploración visual sobre el documento, identificar los elementos incluidos en la base de datos.

4. Revisar y ejecutar las siguientes consultas a modo de ejemplo:

1) Seleccionar la información del país México
for \$x in mondial/country
where \$x/name="Mexico"
return \$x

2) Mostrar los nombres de las provincias de México
for \$x in mondial/country
where \$x/name="Mexico"
return \$x/province/name

3) Seleccionar la capital de México

```
for $x in mondial/country
where $x/name="Mexico"
return $x/@capital
```

```
4) seleccionar los grupos étnicos de México
for $x in mondial/country
where $x/name="Mexico"
return
$x/ethnicgroup
```

```
5) Mostrar el nombre y profundidad de los mares
for $x in mondial/sea
where $x/name
return
<res> { $x/name } { $x/depth }
</res>
```

```
6) Mostrar las información de las montañas de Puebla
for $x in mondial/country[name="Mexico"]/province,
    $y in mondial/mountain
where $x/name="Puebla" and $y/located/@province = $x/@id
return $y
```

5. Escribir las siguientes consultas en XQuery. Usando la base de datos mundial.xml

- 1) Mostrar la capital y área de Canada
- 2) Mostrar lagos de estados unidos
- 3) Mostrar el nombre de las montañas de estaos unidos y la provincia donde están.
- 4) Mostrar todas las montañas de Brasil
- 5) Contar los países que hablan inglés
- 6) Los lagos con profundidad mayor a 50m
- 7) Mostrar el nombre de las ciudades de estados unidos
- 8) Nombre de aeropuertos con elevación mayor a 1000 y nombre del país donde está.
- 9) Población de Manchester Inglaterra
- 10) Mostrar el nombre de los mares (y su profundidad) ordenados ascendentemente por profundidad
- 11) Nombre de los países con más de 5 religiones
- 12) Nombre de los ríos de Brasil
- 13) Nombre de las provincias de la India y su capital
- 14) Mostrar el nombre de los países donde se habla español e inglés
- 15) Mostrar las organizaciones de las que es miembro México

Referencias

- XML, <https://www.w3.org/standards/xml/core>
- XQuery, https://www.w3schools.com/xml/xquery_intro.asp
- Software BaseX, <https://basex.org/download/>
- Base de datos 'Mondial0, <https://www.dbis.informatik.uni-goettingen.de/Mondial/>