

UT05. 01_XML. TRATAMIENTO Y RECUPERACIÓN DE DATOS

Curso: 2DAW.

Lenguajes de Marcas y sistemas de gestión de información

Introducción

Introducción

- Bases de datos XML nativas, alternativa a las relacionales
- BaseX
 - XPath
 - XQuery
 - XLink
 - XPointer
 - SGBD con soporte XML
 - Lenguaje para procesar XML

BD de datos XML nativas

- Almacén de documentos XML
- Ejemplos: BaseX, eXist o Qizx
- BaseX
 - Motor de BD nativo XML
 - Ligero, alto rendimiento en operaciones y fácilmente escalable
 - Incluye procesadores de XPath y XQuery
 - Puede crear una BD constituida por 1 ó más documentos XML y acceder al contenido de los documentos con XPath y XQuery

BaseX

- Vistas → formas de visualización del documento con el que se está trabajando, sincronizadas entre sí
 - Texto plano
 - Mapa
 - Árbol
 - Serie de carpetas
 - Tabla
 - Diagrama de dispersión

BaseX

The screenshot displays the BaseX 7.9 application window titled "BaseX 7.9 - formacionProfesional". The interface is divided into several panes:

- Top Bar:** Contains menu items (Database, Editor, View, Visualization, Options, Help) and a toolbar with icons for file operations and navigation.
- Find Bar:** A search bar labeled "Find..." with a dropdown arrow and a magnifying glass icon.
- Left Pane (File Explorer):** Shows the directory structure of the project located at "C:/Program Files (x86)/BaseX". It lists files such as ".basex (577 Bytes)", ".basexgui (2 KB)", ".basexhome (0 Bytes)", "BaseX.exe (3 MB)", "basex.jar (3 MB)", and "changelog.txt (21 KB)".
- Editor Pane:** Displays the XML document being edited. The content shows two cycles defined by the user:

```
<formacionProfesional>  
  <ciclos>  
    <ciclo siglas="SMR">  
      <nombre>Sistemas microinformáticos y redes</nombre>  
      <familiaProfesional>Informática y comunicaciones</familiaProfesional>  
      <duracion unidad="horas">2000</duracion>  
      <grado>Medio</grado>  
      <referenteInternacional clasificacion="CINE">3</referenteInternacional>  
    </ciclo>  
    <ciclo siglas="ASIR">  
      <nombre>Administrador de sistemas informáticos en red</nombre>  
      <familiaProfesional>Informática y comunicaciones</familiaProfesional>  
      <duracion unidad="horas">2000</duracion>  
      <grado>Superior</grado>  
      <referenteInternacional clasificacion="CINE">5b</referenteInternacional>  
    </ciclo>  
  </ciclos>  
</formacionProfesional>
```
- Right Pane (Query Results):** Titled "0 Results", it shows a visual representation of the XML data as a tree structure. It highlights the "ciclos" element and its children, including "nombre", "familiaProfesional", "duracion", "grado", and "referenteInternacional".
- Bottom Left Pane (Result View):** Displays the raw XML output of the query, which matches the content shown in the editor pane.
- Bottom Right Pane (Query Info):** Provides details about the executed query.
 - Total Time:** 75.27 ms
 - Command:** CREATE DB formacionProfesional "C:/Users/mc/Documents/Portatil/Documents/Curso 14-15/Lenguajes de Marcas/UT5/Ejemplos XPath/formacionProfesional.xml"
 - Result:** Database 'formacionProfesional' created in 73.14 ms.

At the bottom left, a status bar indicates "Time needed: 75.27 ms". At the bottom right, a small indicator shows "41 MB" of memory usage.

BaseX

- Secciones más importantes de la interfaz gráfica de BaseX:
 - Barra de menús en la parte superior.
 - Barra de herramientas (botones) inmediatamente debajo.
 - Línea de comandos justo debajo.
 - Editor de consultas.
 - Diversas visualizaciones de los datos, en concreto de árbol, de mapa y de texto.
 - Información sobre las consultas ejecutadas (Query Info).
- Permite ejecutar tres tipos de sentencias:
 - Opción **Command**: Órdenes propias de cualquier sistema gestor. Existen, comandos propios de un sistema gestor como CREATE DB, OPEN, CREATE índice , CREATE USER, ADD, DELETE, REPLACE...
 - Opción **Search/Find**: Expresiones XPath, un lenguaje de expresiones.
 - Opción **XQuery**: XQuery es un lenguaje de consulta.

Para comenzar...

- Crear bases de datos XML nativas.
 - Seleccionar el menú **Database** y la opción **New**.
 - Seleccionar un archivo XML (**Browse...**) o un directorio que contenga algún archivo XML. BaseX puede crear bases de datos asociadas a uno o a varios archivos XML.
 - Asignar un nombre a la base de datos.
 - Seleccionar lo que se quiera ejecutar, órdenes del sistema gestor (Command), expresiones XPath (Search/Find) o códigos Xquery (XQuery).

XPath

- **XPath + XSLT + XSL-FO** = familia de lenguajes **XSL** (acceder, transformar y dar formato de salida a documentos XML).
- Documento XML → estructura de un árbol
 - Elemento padre
 - Elemento hijo
 - Hermano
 - Antepasado (o ascendiente)
 - Descendiente...
- **XPath** es un lenguaje de expresiones que no se usa de manera independiente, sino que se emplea en el contexto de un lenguaje anfitrión, como XSLT.

XPath. Direccionamiento

- Las expresiones XPath pueden ser
 - Absolutas (empiezan por /)
 - Relativas (con respecto a nodo determinado, llamado nodo de contexto).
- XPath trata un documento XML como un árbol de nodos.

Xpath. Direccionamiento

- Tipos de nodos en XPath
 - **Raíz** del documento (uno por documento): contiene todo el documento y se representa por el símbolo /. No tiene nodo padre y tiene al menos un nodo hijo (nodo documento), comentarios e instrucciones de procesamiento, que no forman parte del documento en sí.
 - **Elemento**: todo elemento de un documento XML es un nodo de XPath.
 - **Atributo**: todo atributo de un documento XML es un nodo de XPath.
 - **Texto**: todo contenido textual de un elemento es un nodo de XPath.
 - **Comentario**.
 - **Instrucciones de procesamiento**.
 - **Espacios de nombres**.
- Los tipos de datos básicos son: string, number, boolean y node-set (conjunto de nodos).

Xpath. Direccionamiento

- Expresiones más comunes

Expresión	Coincidencia
elemento	Elemento de nombre elemento
/elemento	Elemento de nombre elemento ubicado en la raíz del documento
e1/e2	Elemento e2 hijo directo del elemento e1
e1 //e2	Elemento e2 descendiente (hijo, nieto, bisnieto...) del elemento e1
//element	Elemento de nombre elemento ubicado en cualquier nivel por debajo de la raíz del documento
@atributo	Atributo de nombre atributo
*	Cualquier elemento (todos los elementos)
@*	Cualquier atributo (todos los atributos)
.	Nodo actual
..	Nodo padre
espNom:*	Todos los elementos en el espacio de nombres de prefijo espNom
@espNom:*	Todos los atributos en el espacio de nombres de prefijo espNom

XPath.

- Ejemplo *formacionProfesional.xml*

Xpath. Acceso a...

- Elementos/atributos vs. su contenido textual
 - Una ruta en XPath devuelve el conjunto de nodos descendientes. Si no existiera esa ubicación, se devolvería el conjunto vacío.
 - Función **text ()** a continuación de la expresión al nodo: devuelve su contenido textual.
 - Acceso a un atributo: **@nombre-atributo**
 - Función **data()** a continuación del atributo: devuelve su valor (en algunos analizadores no funciona)
 - En XPath 2.0 se usa la misma función **string()** para ambos casos:

Xpath. Acceso a...

- Elementos con **rutas simples**: direccionamiento absoluto.
 - Todos los elementos hijos del nodo raíz, es decir, el elemento documento.
 - Todos los atributos de cualquier elemento del documento XML.
 - Nombre de ciclos formativos.
 - Duración de los módulos profesionales.
 - Elementos que se llamen *nombre* ubicados en cualquier lugar del documento.
 - Siglas de los ciclos formativos.
 - Todos los atributos de los módulos.

Xpath. Elementos del lenguaje

- Operadores

- Operadores **booleanos**

- **and**
 - **or**
 - **not**
 - **=**
 - **!=**
 - **>** (>)
 - **<** (<)
 - **>=** (>=)
 - **<=** (<=)

- Operadores **aritméticos**

- **+** Suma
 - **-** Resta
 - ***** Multiplicación
 - **div** División
 - **mod** Resto (módulo)

- **Otros** operadores

- **|** Unión de resultados

Xpath. Elementos del lenguaje

- 8. Descendientes de /formacionProfesional , de nombre ciclo o módulo, que contengan un descendiente directo nombre.

Xpath. Elementos del lenguaje

- Funciones, entre corchetes [**funcion()**]
 - Funciones **numéricas**
 - **round()** --Redondeo --**round(3.14) = 3**
 - **abs()** --Valor absoluto --**abs(-7) = 7**
 - **floor()** --Suelo --**floor(7.3) = 7**
 - **ceiling()** --Techo -- **ceiling(7.3) = 8**
 - Funciones de **cadena**
 - **substring()** --Subcadena --**substring('Beatles', 1,4) = 'Beat'**
 - **starts-with()** --Cadena comienza por --**starts-with('XML','X') = true**
 - **contains()** --Cadena contiene --**contains('XML','XM') = true**
 - **normalize-space()** --Espacios normalizados --**normalize-space(' The end ') = 'The end'**
 - **translate()** --Cambia caracteres en una cadena --**translate('The end','The','An') = 'An end'**
 - **string-length()** --Longitud de una cadena --**string-length('Beatles') = 7**

Xpath. Elementos del lenguaje

- 9. Nombre de los ciclos cuyas siglas empiezan por D.

Xpath. Elementos del lenguaje

- Funciones que devuelven elementos por su posición
 - **position()=n** --Elemento que se encuentra en la n-ésima posición
 - **<elemento>[n]** --Elemento que se encuentra en la n-ésima posición
 - **last()** --El último elemento de un conjunto
 - **last() - i** --El último elemento - i (ej. Si $i=1 \rightarrow$ el penúltimo)
- 10. Todos los datos del segundo ciclo
- 11. Todos los datos del último ciclo

Xpath. Elementos del lenguaje

- Funciones que devuelven nodos
 - **name ()** --Nombre del nodo actual
 - **root ()** --El elemento raíz
 - **node ()** --Nodos descendientes del actual
 - **comment()** --Comentarios
 - **processing-instruction()** --Instrucciones de procesamiento
- 12. Devuelve el nombre del nodo actual, es decir, ciclo
- 13. Todos los comentarios existentes en el documento
- 14. Todas las instrucciones de procesamiento existentes en el documento
- 15. Todos los nodos del árbol cuyo nombre tenga una longitud de cuatro caracteres

Xpath. Elementos del lenguaje

- Funciones de **agregado**
 - **count()** Conteo de elementos
 - **avg ()** Media de valores
 - **max ()** Valor máximo
 - **min ()** Valor mínimo
 - **sum ()** Suma de valores
- 16. Nombre de los módulos que se cursan en dos ciclos

Xpath. Acceso a elementos con filtros de valores literales

- Extensión de las expresiones anteriores mediante la cual se puede seleccionar elementos en función del valor de sus atributos o del propio elemento.
 - 17. Datos del módulo de código 0373
 - 18. Nombre de los ciclos de grado medio
 - 19. Duración de los módulos impartidos en el DAW
 - 20. Nombre de los ciclos que tengan un reconocimiento de ECTSs

Xpath. Acceso a elementos con filtros de valores literales

- Nombre de los ciclos cuyas siglas son ASIR o SMR
- Nombre de los ciclos cuyas siglas se encuentren, en orden alfabético, entre ASIR y SMR
- Datos del segundo modulo de la lista de los cursados en primer curso
- Datos del tercer ciclo que tenga reconocimiento en ECTSs
- Datos de los módulos en las posiciones 3 a 5 (sólo en XPath 2.0)
- Nombre de los módulos cuya relación (cociente) entre duración y ects sea mayor o igual de 10

Xpath. Acceso a elementos con filtros de valores recuperados

- Nombre de los ciclos en los que se cursen módulos de una duración de 80 horas o más.
- Nombres de módulos que tengan una duración mayor que la de Sistemas Operativos monopuesto. Se usa la función **number()** para obtener el valor numérico de la duración, de lo contrario haría la comparación como cadenas (donde se cumple que “100” < “20”).

Xpath. Acceso a elementos mediante ejes

- **Ejes** → expresiones que permiten acceder a trozos del árbol XML, apoyándose en las relaciones de parentesco entre los nodos.
 - **self::*** Devuelve el propio nodo de contexto. Equivalente al punto .
 - **child::*** Devuelve los nodos “hijo” del nodo de contexto.
 - **parent::*** Devuelve el nodo padre del nodo contexto. Equivale a ..
 - **ancestor::*** Devuelve los “antepasados” (padre, abuelo, hasta el nodo raíz) del nodo de contexto
 - **ancestor-or-self::*** Devuelve los nodos “antepasados” del nodo de contexto además del propio nodo de contexto.
 - **descendant::*** Devuelve los nodos “descendientes” (hijo, nieto...) del nodo de contexto.

Xpath. Acceso a elementos mediante ejes

- **descendant-or-self::*** Devuelve los nodos “descendientes” (hijo, nieto...) del nodo de contexto además del propio nodo de contexto. Equivalente a //
- **following::*** Devuelve los nodos que aparezcan después del nodo de contexto en el documento, excluyendo a los nodos descendientes, los atributos y los nodos de espacio de nombres.
- **preceding::*** Devuelve los nodos que aparezcan antes del nodo de contexto en el documento, excluyendo a los nodos ascendientes, los atributos y los nodos de espacio de nombres.
- **preceding-sibling::*** Devuelve los “hermanos mayores” del nodo de contexto.
- **following-sibling::*** Devuelve los “hermanos menores” del nodo de contexto.
- **attribute::*** Atributos del nodo contexto. Equivale a @namespace:: * Espacio de nombres del nodo de contexto.

Xpath. Acceso a elementos mediante ejes

- Elementos hermanos menores del segundo módulo.
- Descendientes del elemento raíz que se llamen nombre.
- ¿Qué visualiza **/formacionProfesional/ciclos/ciclo?**
- ¿Y **/formacionProfesional/ciclos/ciclo/child::*?**

Ejemplo	Resultado
child::book	Selecciona todos los nodos libros que son hijos del nodo actual
attribute::lang	Selecciona el lenguaje del nodo actual
child::*	Selecciona todos los elementos hijos del nodo actual
attribute::*	Selecciona todos los atributos del nodo actual
child::text()	Selecciona todos los nodos texto del nodo actual
child::node()	Selecciona todos los nodos hijos del nodo actual
descendant::book	Selecciona todos los descendientes del nodo actual
ancestor::book	Selecciona todos los nodos ascendientes del nodo actual
ancestor-or-self::book	Selecciona todos los nodos ascendientes libro del nodo actual, incluido el nodo actual
child::* / child::price	Selecciona todos los abuelos price del nodo actual