XML PATH (Junio 2014)

Raúl Andrés Ortiz Fuentes, 112508.

*Índice de términos*—XSLT: es un estándar de la organización W3C que presenta una forma de transformar documentos XML en otros e incluso a formatos que no son XML. Grafo Dirigido: Un grafo dirigido es un grafo en el que los arcos tienen dirección.

# INTRODUCCIÓN

E

n el siguiente articulo se hablara sobre el lenguaje XPath, lenguaje que permite la lectura y procesamiento de cualquier archivo XML, a través del uso de expresiones regulares que permitirán la selección de nodos, atributos y elementos.

# XML PATH

Es un lenguaje que permite seleccionar nodos de un documento XML y calcular valores a partir de su contenido. La selección del contenido se realiza a través de expresiones regulares que recorren y procesan el XML. La idea es seleccionar partes de un texto sin atributo, lo que permite realizar búsquedas y seleccionar teniendo en cuenta la estructura jerárquica del XML. XPath fue creado para su uso en el estándar XSLT, en el que se usa para seleccionar y examinar la estructura del documento de entrada de la transformación.

# Árbol del documento

Un documento XML es procesado por un analizador construyendo un árbol de nodos. Este árbol comienza con un elemento raíz, que se diversifica a lo largo de los elementos que cuelgan de él y acaba en nodos hoja, que contienen solo texto, comentarios, instrucciones de proceso o incluso que están vacíos y solo tienen atributos.

La forma en que XPath selecciona partes del documento XML se basa precisamente en la representación arbórea que se genera del documento. De hecho, los "operadores" de que consta este lenguaje nos recordarán la terminología que se utiliza a la hora de hablar de árboles en informática: raíz, hijo, ancestro, descendiente, etc.

Un caso especial de nodo son los nodos atributo. Un nodo puede tener tantos atributos como desee, y para cada uno se le creará un nodo atributo. No obstante, dichos nodos atributo no se consideran como hijos suyos, sino más bien como etiquetas añadidas al nodo elemento.

## 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:

1. *Nodo Raíz*

Es el origen de todo, de aquí nacen los demás nodos y se identifica con el símbolo “/”. No se debe confundir con el elemento raíz del documento XML.

1. *Nodo Elemento*

Cualquier elemento de un documento XML se convierte en un nodo elemento dentro del árbol. Cada elemento tiene su nodo padre. El nodo padre de cualquier elemento es, a su vez, un elemento, excepto el elemento raíz, cuyo padre es el nodo raíz. Los nodos elemento tienen a su vez hijos, que son: nodos elemento, nodos texto, nodos comentario y nodos de instrucciones de proceso. Los nodos elemento también tienen propiedades tales como su nombre, sus atributos e información sobre los "espacios de nombre" que tiene activos.

1. *Nodos texto*

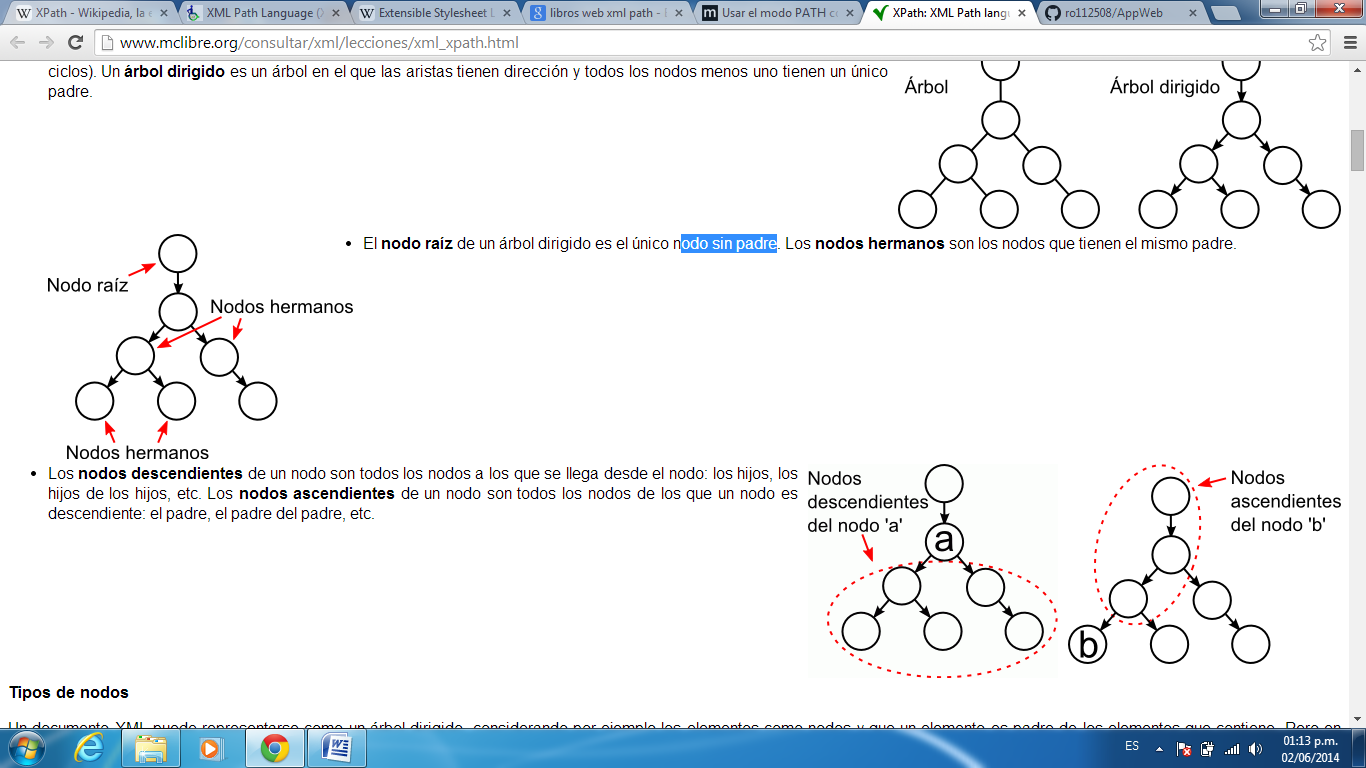
Por texto se hace referencia a todos los caracteres del documento que no están marcados con alguna etiqueta. Un nodo texto no tiene hijos, es decir, los distintos caracteres que lo forman no se consideran hijos suyos.

1. *Nodos atributo*

Los nodos atributo no son tanto hijos del nodo elemento que los contiene como etiquetas añadidas a dicho nodo elemento. Cada nodo atributo consta de un nombre, un valor y un posible "espacio de nombres".

1. *Nodos comentario y de instrucciones de proceso*

Aparte de los nodos indicados, en el árbol también se generan nodos para cada nodo con comentarios y con instrucciones de proceso. Al contenido de estos nodos se puede acceder con la propiedad *string-value*.

  
Fig. 1 Representación de un XML a través de un grafo dirigido.

En la Fig.1 se da una representación visual del árbol que se puede extraer de un documento XML, el cual está compuesto por el nodo raíz y los nodos hijos que a su vez se convierte en padres de otros nodos formado nodos descendientes (hijos de los hijos) y ascendientes (padre del padre).

# 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:

1. *Sintaxis Abreviada*: más compacta y fácil de leer
2. *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:

1. *Eje*: indica el nodo o los nodos en los que se realiza la búsqueda. Se define como la relación existente entre un paso de localización y su nodo contextual. También se puede definir como un conjunto de nodos relativos al nodo contextual.
2. *Nodo de comprobación:* especifica el nodo o los nodos seleccionados dentro del eje
3. *Predicado*: permite restringir los nodos de comprobación. Un predicado es una "condición" que permiten seleccionar un nodo con unos determinados atributos o características.

## Componentes de la Sintaxis:

A continuación se mostraran los elementos con los que se construyen las expresiones XPath para el recorrido y el procesamiento de un documento XML. Las expresiones serán dividas en las tres partes indicadas anteriormente *eje, nodo de comprobación y predicado.*

1. *Eje:*

* **“/”:** 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.

1. *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**

1. *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.

# CONCLUSIONES

En conclusión, XPath es un lenguaje de "ayuda" para localizar elementos dentro de cualquier fichero XML, con este se puede realizar la comprobación del contenido y a su vez de su estructura, ya que nos permite la lectura y procesamiento del contenido de forma rápida y sin mayor complejidad.

Referencias

1. Bartolomé Marco, “XPath: XML Path language”, 2014, Recuperado de: <http://www.mclibre.org/consultar/xml/lecciones/xml_xpath.html>.
2. Madrid Víctor, “Introducción teórica a XPath”, 2009, Recuperado de: <http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=teoriaXPath#04>.

**Primer autor** Raúl Andrés Ortiz Fuentes, 112508, Aplicaciones Web Avanzadas.