

Lenguaje XPATH

Jara Rodríguez Carlos Eduardo.
cj102509@ingenieria.sanmartin.edu.co
Fundación Universitaria San Martín.

Resumen— es un lenguaje que permite construir expresiones que recorren y procesan un documento XML. La idea es parecida a las expresiones regulares para seleccionar partes de un texto sin atributos (plain text). XPath permite buscar 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.

I. INTRODUCCIÓN

Todo el procesamiento realizado con un fichero XML está basado en la posibilidad de direccionar o acceder a cada una de las partes que lo componen, de modo que podamos tratar cada uno de los elementos de forma diferenciada.

El tratamiento del fichero XML comienza por la localización del mismo a lo largo del conjunto de documentos existentes en el mundo. Para llevar a cabo esta localización de forma unívoca, se utilizan los URI (Uniform Resource Identifiers), de los cuales los URL (Uniform Resource Locators) son sin duda los más conocidos.

Una vez localizado el documento XML, la forma de seleccionar información dentro de él es mediante el uso de XPath, que es la abreviación de lo que se conoce como XML Path Language. Con XPath podremos seleccionar y hacer referencia a texto, elementos, atributos y cualquier otra información contenida dentro de un fichero XML.

XPath en sí es un lenguaje sofisticado y complejo, pero distinto de los lenguajes procedurales que solemos usar (C, C++, Basic, Java...). Además, como casi todo en el mundo de XML, aún está en estado de desarrollo, por lo que no es fácil encontrar herramientas que incorporen todas sus

funcionalidades.

XPath es a su vez la base sobre la que se han especificado nuevas herramientas que aprovechan para el tratamiento de documentos XML. Herramientas tales como XPointer, XLink y XQuery (el lenguaje que maneja los documentos XML como si de una base de datos se tratase). Así, XPath sirve para decir cómo una hoja de estilo debe procesar el contenido de una página XML, pero también para poder poner enlaces o cargar en un navegador zonas determinadas de una página XML, en vez de toda la página.

XPath es el resultado de un esfuerzo por proporcionar una sintaxis y la semántica para la funcionalidad compartida entre Transformaciones XSL común [XSLT] y XPointer [XPointer]. El propósito principal de XPath es para hacer frente a partes de un XML [XML] documento. En apoyo de este objetivo principal, sino que también proporciona servicios básicos para la manipulación de cadenas, números y booleanos. XPath utiliza una sintaxis compacta, que no es XML para facilitar el uso de XPath dentro de los URI y los valores de atributos XML. XPath opera sobre la estructura lógica abstracta de un documento XML, en lugar de su sintaxis superficial. XPath obtiene su nombre de su uso de una notación de ruta como en las direcciones URL para navegar a través de la estructura jerárquica de un documento XML.

Además de su uso para el direccionamiento, XPath también está diseñado de manera que tiene un subconjunto natural que puede ser utilizado para hacer coincidir (probar si o no un nodo coincide con un patrón); este uso de XPath se describe en XSLT.

Modelos XPath un documento XML como un árbol de nodos. Hay diferentes tipos de nodos, incluyendo nodos elemento, nodos atributo y nodos de texto. XPath define una forma de calcular un valor de cadena para cada tipo de nodo. Algunos tipos de nodos también tienen nombres. XPath es totalmente compatible con los espacios de nombres XML [XML Names]. Por lo tanto, el nombre de un nodo se modela como un par que consiste en una parte local y, posiblemente, un espacio de nombres URI nulo; esto se llama un nombre expandido. El modelo de datos se describe en detalle en [5 Modelo de Datos].

La construcción sintáctica primaria en XPath es la expresión. Una expresión coincide con la producción Expr. Una expresión se evalúa para producir un objeto, que tiene uno de los siguientes cuatro tipos básicos:

node-set (una colección desordenada de nodos sin duplicados)
 boolean (verdadero o falso)
 número (un número de coma flotante)
 cadena (una secuencia de caracteres UCS)
 La evaluación de expresiones se produce con respecto a un contexto. XSLT y XPath especifican cómo se determina el contexto de las expresiones XPath usadas en XSLT y XPath respectivamente. El contexto consiste en:

un nodo (el nodo de contexto)
 un par de no-cero enteros positivos (la posición de contexto y el tamaño de contexto)
 un conjunto de asignaciones de variables
 una biblioteca de funciones
 el conjunto de las declaraciones de espacio de nombres en el ámbito de la expresión
 La posición de contexto es siempre menor o igual al tamaño de contexto.

EJEMPLOS

La sintaxis básica de XPath es similar a la del direccionamiento de archivos. Un camino que se inicia con '/', representa un camino absoluto hacia el elemento requerido.

/AAA

Selecciona el elemento raíz AAA

```
<AAA>
  <BBB/>
  <CCC/>
  <BBB/>
  <BBB/>
  <DDD>
    <BBB/>
  </DDD>
  <CCC/>
</AAA>
```

/AAA/CCC

Selecciona todos los elementos CCC que son hijos del elemento raíz AAA

```
<AAA>
  <BBB/>
  <CCC/>
  <BBB/>
  <BBB/>
  <DDD>
    <BBB/>
  </DDD>
  <CCC/>
</AAA>
```

REFERENCIAS

- [1] http://zvon.org/xxl/XPathTutorial/Output_spa/examples.html
- [2] http://zvon.org/xxl/XPathTutorial/Output_spa/example1.html
- [3] http://www.programacion.com/articulo/xslt-_xslt-_xpath:_que_es_cada_cosa_1

Autor

Carlos Eduardo Jara Rodríguez
 Estudiante de ingeniería de sistemas VIII semestre
 Fundación Universitaria San Martín