

XML PATH

López, Juan.
jl091505@ingenieria.sanmartin.edu.co
Fundación Universitaria San Martin

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.

Índice de Términos— XML, Path, XPAth

I. USO XPATH.

A. Xpath

El objetivo de XPath es la generación de expresiones o mejor dicho, expresiones de trayectorias que navegan por un documento XML (recordar que este tipo de documentos disponen de una estructura jerárquica), para ello modela dicho documento como un árbol de nodos. También incorpora una biblioteca con funciones estándar.

Ayuda : Sería algo similar a seleccionar partes de un texto plano mediante el uso de expresiones regulares.

XPath ha servido de base para el desarrollo de múltiples herramientas para el tratamiento de documentos XML.

1. XLink
2. XPointer
3. XQL

B. XLink

XLink (Lenguaje de Enlace XML) es un lenguaje que permite generar enlaces en los ficheros XML, permitiendo establecer relaciones cruzadas (enlaces) entre diferentes elementos (2 o más) sin que estos sepan que están enlazados.

Hay que diferenciar 2 tipos de enlaces:

Simples: Son los enlaces de un recurso local a uno remoto.

Extendidos: Son los enlaces que permiten vincular muchos recursos entre sí.

También es una recomendación de W3C.

C. Xpointer

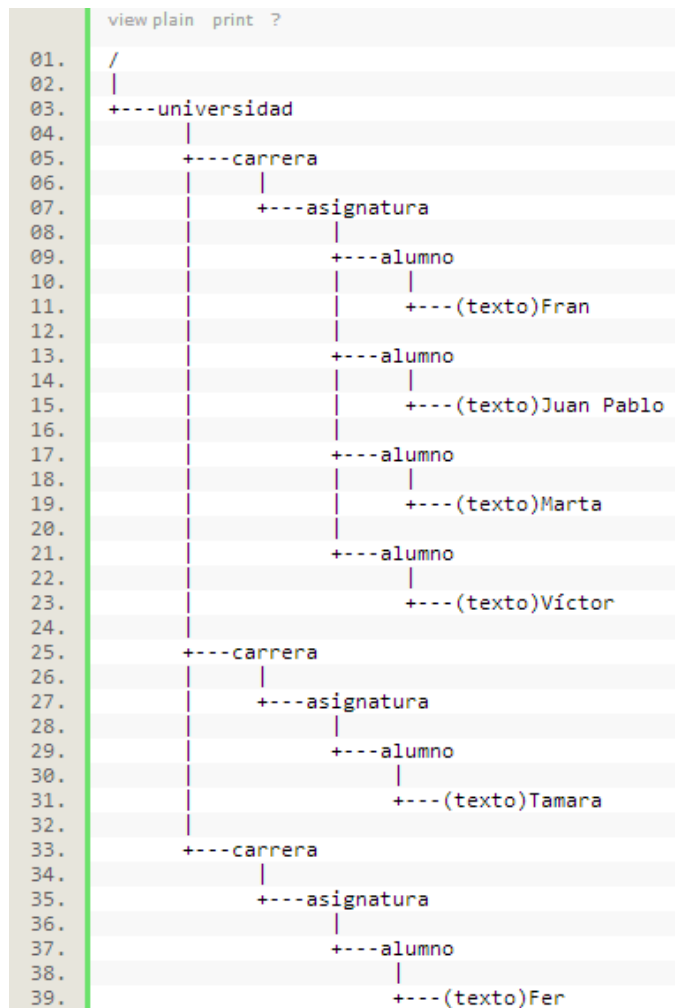
XPointer (Lenguaje de Direccionamiento XML) es un lenguaje que permite identificar unívocamente partes de un documento XML para establecer vínculos. Lo que hace es establecer un tipo de direccionamiento del documento XML en base a su estructura interna (elementos, atributos, etc.). Entre sus aplicaciones estaría el mostrar partes concretas de un documento XML.

D. XQL

XQL (Lenguaje de consulta XML) es un lenguaje que permite manejar los documentos XML como si fueran base de datos, es decir, "SQL" para los documentos XML.

E. Construcción árbol de Nodos

Un árbol de nodos se genera tras procesar un documento XML mediante un parser o analizador. Ejemplo del árbol de nodos referente al ejemplo del comienzo de este apartado:



Este árbol de nodos tiene una estructura básica que es común a todos ellos:

- Comienza por un elemento raíz.
- Despliega una serie de elementos intermedios que dependen del elemento raíz.
- Finaliza en varios elementos finales (también denominados nodos hoja).

Nota: El funcionamiento de XPath se basa en este tipo de representación del documento.

Nodo actual (Current)

Es el nodo que está seleccionado cuando se evalúa una expresión XPath, es decir, es el comienzo desde donde se evalúa dicha expresión.

En el ejemplo, si el nodo actual fuera el nodo alumno con texto "Fer" y se buscaran el resto de alumnos, se perderían los que son anteriores.

Nodo contexto (Context)

Son aquellos nodos que son evaluados de forma parcial, para obtener el resultado de la evaluación de la expresión XPath. Hay que tener en cuenta que en cada evaluación de las subexpresiones se obtienen un nuevo conjunto de nodos que pasa a ser el nuevo contexto a evaluar en las siguientes subexpresiones.

Tamaño del contexto

Es el número de nodos que se están evaluando en la expresión XPath en un momento dado.

También se denomina tamaño contextual.

Nota: Siempre se corresponde con un entero positivo (no nulo).

Posición del contexto

Es la posición que ocupa dentro del documento el nodo contexto.

También se denomina posición contextual.

Tipos de Nodo

Estos son los diferentes tipos de nodos que podemos distinguir:

Nodo raíz (Root)

Es común a cualquier árbol de nodos.

Se reconoce por su identificador "/".

Es muy importante no confundirlo con el elemento raíz del documento. Por lo tanto y aplicado a nuestro ejemplo:

Nodo raíz : es el elemento "/" del árbol.

Elemento raíz : es el elemento "universidad".

Se puede observar que el elemento raíz está contenido en el nodo raíz

Nodo elemento (Element)

Cualquier elemento del documento es un nodo elemento del árbol de nodos.

Características:

- Cada elemento (nodo elemento) tiene un nodo padre.
- El nodo padre de un elemento es también un elemento excepto el padre de elemento raíz que es el nodo raíz.
- Cada nodo elemento tiene hijos que pueden ser otros nodos elemento o nodos hoja.
- Identificador único (Si se acompaña de un DTD que especifique dicho atributo).
- Disponen de propiedades (nombre/atributos/etc.)

En nuestro ejemplo, si seleccionamos el elemento "asignatura"

Nodo padre : "carrera".

Nodo/s hijo/s: los 6 nodos elemento "alumnos".

Nodo atributo (Attribute)

Es un tipo especial de nodo.

Características:

- Un nodo no tienen un número determinado de atributos.
- Atributo: es una etiqueta incorporada al elemento que lo contiene. (Se compone de un nombre y un valor (Formato cadena).
- Los atributos pueden tener valores por defecto si se especifican en el DTD.

```

01. <universidad>
02.   <carrera nombre="Ing. Informática">
03.     <asignatura>
04.       <alumno>Fran</alumno>
05.       <alumno>Juan Pablo</alumno>
06.       <alumno>Marta</alumno>
07.       <alumno>Víctor</alumno>
08.     </asignatura>
09.   </carrera>
10.   <carrera nombre="Económicas">
11.     <asignatura>
12.       <alumno>Tamara</alumno>
13.     </asignatura>
14.   </carrera>
15.   <carrera nombre="Ing. Industriales">
16.     <asignatura>
17.       <alumno>Fer</alumno>
18.     </asignatura>
19.   </carrera>
20. </universidad>

```

Expresiones XPath.

En XPath una expresión es el elemento base que se utiliza para producir un resultado.

Estableciendo una similitud con cualquier lenguaje de programación, una expresión vendría a ser una instrucción (recordar que XPath es un lenguaje declarativo).

Mediante las expresiones se van a poder seleccionar nodos o conjuntos de nodos.

Al evaluar una o varias expresiones se genera un resultado, este resultado podrá ser de uno de los siguientes tipos:

- **Conjunto de nodos (Node-set)** : Un conjunto de nodos (Nota :sin duplicados y sin orden).
- **Cadena** : Un conjunto de caracteres.
- **Número** : Punto flotante.
- **Booleano** : Verdadero / Falso.

Localización (Location Path).

Una ruta de localización (location path) es uno de los tipos de expresiones más importantes que se pueden especificar en XPath, el resultado generado siempre es un node-set.

Importante: Se devuelven los nodos considerados, no los hijos de los nodos.

Su sintaxis es muy similar a la usada para describir la ruta de los ficheros en Linux o Unix. También se pueden indicar que serían similares a las rutas de Windows y MS-DOS sin consideramos 2 excepciones:

- Omitir la unidad de disco. Por ejemplo : "C:".
- Modificar las barras "\" por barras "/".

Pasos de localización

Paso de localización es el nombre con el que se denomina a cada cambio que se produce en una ruta de localización. Es decir, a cada cambio de "/" que se produce.

Un paso de localización se compone de:

- Eje (Axis) : Es la relación entre el nodo contextual y el paso.
- Nodo prueba (Node test) : Es el elemento filtrado (En el caso de rutas de directorios sería el nombre del directorio).
- Predicado (Predicate) : Expresión XPath entre corchetes.

El Eje a veces está implícito.

El nodo prueba se suele identificar por el nombre, aunque poniendo "*" simboliza cualquier nombre.

El predicado es opcional.

Predicados

Un predicado es una "condición" que permiten seleccionar un nodo con unos determinados atributos o características. Como se ha visto en puntos anteriores incorpora un tipo de verificación durante el paso de localización.

Su sintaxis viene dada entre corchetes ([predicado]) y tiene un valor booleano.

Ejemplo de predicado:

```
/universidad/carrera/asignatura[@nombre="Base de datos"]/alumno
```

Hace referencia a todos los alumnos que pertenecen a las asignaturas que tiene como atributo nombre "Base de datos".

En el apartado anterior se dijo que los predicados son opciones y que en caso de existir se representan entre corchetes. Entre los corchetes se pueden colocar

Ejes

Un eje 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.

Recordar la sintaxis de un paso de localización:

Eje::NodoPrueba[Predicado]

Como se puede observar

El eje y el nodo de prueba se encuentran separados por el elemento "::".

Se puede traducir por : NodoPrueba[Predicado] "que es un " Eje.

Funciones

En XPath esta permitido el uso de funciones que nos van a facilitar poder realizar algunas operaciones necesarias de una manera sencilla.

Podemos encontrar una amplia variedad de funciones como por ejemplo:

- boolean() : Convierte a booleano. Al aplicarlo sobre un conjunto de nodos devuelve true si esta vacío.
- position() : Devuelve la posición de un nodo en su contexto
- last() : Devuelve la última posición
- count : Devuelve el n.º de nodos e un conjunto de nodos.
-
- not() : Devuelve el contrario de un booleano dado.

Conclusión

En resumen, XPath es un lenguaje de "ayuda" para localizar elementos dentro de cualquier fichero XML, como véis tiene muchas utilidades pero seguro que vosotros sois capaces de sacarles muchas más. Tenerlo muy en cuenta cuando uséis el paquete Selenium ya que os facilitará en gran medida el

trabajo y sobre todo permitirá que no se produzcan errores en los scripts cuando estos se basan en la posición de los elementos.

REFERENCIAS

- [1] <http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=teoriaXPath>