

Andy Asael Sánchez Jiménez

201904024

XML – DOM

El Document Object Model (DOM) es la base de XML. Los documentos XML tienen una jerarquía de unidades llamadas nodos; DOM, es una forma de describir los nodos y las relaciones entre ellos.

Un documento DOM es una colección de nodos o piezas de información organizada en una jerarquía. Esta jerarquía permite a los desarrolladores para navegar por el árbol para buscar información específica. Porque se basa en una jerarquía de información, de la catedral se dice que es base de árbol.

El XML DOM, por otro lado, también proporciona una API que permite a los desarrolladores agregar, editar, mover o eliminar los nodos en el árbol en cualquier momento con el fin de crear una aplicación.

El responsable del DOM es el World Wide Web Consortium (W3C). El DOM permite el acceso dinámico a través de la programación para acceder, añadir y cambiar dinámicamente contenido estructurado en documentos con lenguajes como ECMAScript (JavaScript).

El DOM define la manera en que objetos y elementos se relacionan entre sí en el navegador y en el documento.

Puede utilizarse cualquier lenguaje de programación adecuado para el diseño web. En el caso de JavaScript, cada objeto tiene un nombre, el cual es exclusivo y único. Cuando existe más de un objeto del mismo tipo en un documento web, estos se organizan en un vector.

Es posible asignarle una identificación a un objeto, y luego usarla para hacer referencia a este, por ejemplo:

```
<div id="Juan">....</div>
```

Para hacer referencia a elementos del mismo tipo, los cuales, como se ha dicho, están organizados en un vector, se pueden utilizar la notación punto de la siguiente manera.

```
document.div[0]
```

```
document.div["Juan"]
```

```
document.div.Juan
```

Donde el elemento «Juan» es el primero del vector de elementos del tipo <div>. También se puede usar la función getElementById.

Xpath (módulo python)

Xpath es un módulo que es parte de la librería xml.etree.ElementTree por lo general la misma se importa de la siguiente manera:

```
import xml.etree.ElementTree as ET
```

Xpath provee una serie de expresiones para localizar elementos en un árbol, su finalidad es proporcionar un conjunto de sintaxis, por lo que debido a su limitado alcance no se considera un motor en si mismo.

Nos permite extraer facilmente partes del xml haciendo referencia a su ubicación nodal representada a forma de path, lo que nos hace una sintaxis familiarmente sencilla a la hora de construir un parser xml.

Xpath Sintaxis:

tag

Selecciona todos los elementos hijos contenidos en la etiqueta "tag", Por ejemplo: spam, selecciona todos los elementos hijos de la etiqueta spam y así sucesivamente en un path de nodos spam/egg, /spam/egg/milk

*

Selecciona todos los elementos hijos. Ejemplo: */egg, seleccionara todos los elementos nietos bajo la etiqueta egg

.

Selecciona el nodo actual, este es muy usado en el inicio del path, para indicar que es un path relativo.

//

Selecciona todos los sub elementos de todos los niveles bajo el nodo expresado. Por ejemplo: ./egg selecciona todos los elementos bajo egg a través de todo el arbol bajo la etiqueta

..

Selecciona el elemento padre

[@attrib]

Selecciona todos los elementos que contienen el atributo tras el "@"

[@attrib='value']

Seleccione todos los elementos para los cuales el atributo dado tenga un valor dado, el valor no puede contener comillas

[tag]

Selecciona todos los elementos que contienen una etiqueta hijo llamada tag. Solo los hijos inmediatos son admitidos

[tag='text']

Selecciona todos los elementos que tienen una etiqueta hijo llamada tag incluyendo descendientes que sean igual al texto dado

[position]

Selecciona todos los elementos que se encuentran en la posición dada. La posición puede contener un entero siendo 1 la primera posición, la expresión last() para la ultima, o la posición relativa con respecto a la ultima posición last()-1