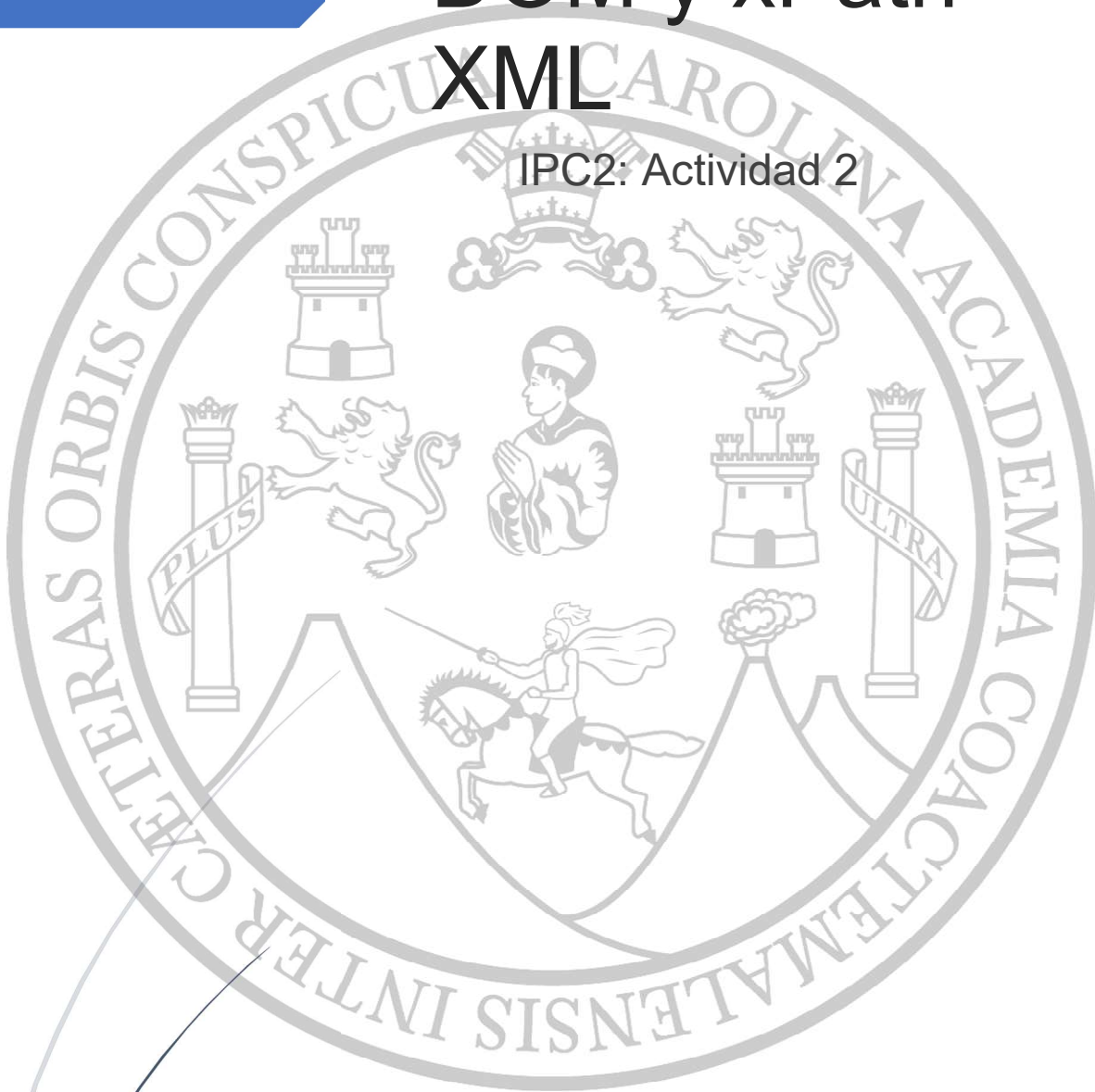


6-2-2021

# DOM y xPath

## XML

IPC2: Actividad 2



Geovanni Nufio - 201901444

USAC

## CONTENIDO

|  |   |
|--|---|
| <b>DOM xml</b> .....                       | 2 |
| <b>¿QUE ES?</b> .....                      | 2 |
| <b>Objetos en el DOM</b> .....             | 2 |
| <b>Objetos de DOMImplementation</b> .....  | 3 |
| <b>Objetos de Nodo</b> .....               | 3 |
| <b>Objetos NodeList</b> .....              | 3 |
| <b>Objetos DocumentType</b> .....          | 4 |
| <b>Objetos Documento</b> .....             | 4 |
| <b>Objetos Elemento</b> .....              | 4 |
| <b>Objetos Atributo</b> .....              | 5 |
| <b>Objetos NamedNodeMap</b> .....          | 5 |
| <b>Objetos comentario</b> .....            | 5 |
| <b>Objetos Texto y CDATASection</b> .....  | 5 |
| <b>Objetos ProcessingInstruction</b> ..... | 5 |
| <b>xpath xml</b> .....                     | 6 |
| <b>¿QUE ES?</b> .....                      | 6 |
| <b>Sintaxis</b> .....                      | 6 |

## ¿QUE ES?

El modelo de Objetos del Documento o por sus siglas “DOM”, este es un API con el cual es posible acceder y modificar archivos XML. Esto es posible gracias a que el DOM presenta los documentos XML como un árbol, esto le permite al cliente construir dichas estructuras desde cero para luego darles acceso a la estructura a través de un conjunto de objetos que implementaron interfaces conocidas. El DOM una vez obtenga un objeto del documento, se puede acceder a las partes del documento XML a través de sus propiedades y métodos. Estas propiedades están definidas en la especificación del DOM.

## OBJETOS EN EL DOM

| Interfaz                     | Sección                       | Propósito  |
|------------------------------|-------------------------------|--|
| <b>DOMImplementation</b>     | Objetos DOMImplementation     | Interfaz para las implementaciones subyacentes                                 |
| <b>Node</b>                  | Objetos Nodo                  | Interfaz base para la mayoría de los objetos en un documento                   |
| <b>NodeList</b>              | Objetos NodeList              | Interfaz para una secuencia de nodos   |
| <b>DocumentType</b>          | Objetos DocumentType          | Información acerca de las declaraciones necesarias para procesar un documento. |
| <b>Document</b>              | Objetos Document              | Objeto que representa un documento entero.                                     |
| <b>Element</b>               | Objetos Elemento              | Nodos elemento en la jerarquía del documento.                                  |
| <b>Attr</b>                  | Objetos atributo              | Nodos de los valores de los atributos en los elementos Nodo.                   |
| <b>Comment</b>               | Objetos Comentario            | Representación de los comentarios en el documento fuente.                      |
| <b>Text</b>                  | Objetos Texto y CDATASection  | Nodos con contenido textual del documento.                                     |
| <b>ProcessingInstruction</b> | Objetos ProcessingInstruction | Representación de instrucción del procesamiento.                               |

## OBJETOS DE DOMIMPLEMENTATION

- DOMImplementation.hasFeature(feature, version)
- DOMImplementation.createDocument(namespaceUri,qualifiedName,doctype)
- DOMImplementation.createDocumentType(qualifiedName,pblicId)

## OBJETOS DE NODO

- Node.nodeType
- Node.parentNode
- Node.attributes
- Node.previousSibling
- Node.nextSibling
- Node.childNodes
- Node.firstChild
- Node.lastChild
- Node.localName
- Node.prefix
- Node.namespaceURI
- Node.nodeName
- Node.nodeValue
- Node.hasAttributes
- Node.hasChildNodes
- Node.isSameNode
- Node.appendChild
- Node.insertBefore
- Node.removeChild
- Node.replaceChild
- Node.normalize
- Node.cloneNode

## OBJETOS NODLIST

- NodeList.item
- NodeList.lenght

## OBJETOS DOCUMENTTYPE

- `DocumentType.publicId`
- `DocumentType.systemId`
- `DocumentType.internalSubset`
- `DocumentType.name`
- `DocumentType.entities`
- `DocumentType.notations`

## OBJETOS DOCUMENTO

- `Document.documentElement`
- `Document.createElement(Nombre)`
- `Document.createElementNS(namespaceURI, tagName)`
- `Document.createTextNode(data)`
- `Document.createComment(data)`
- `Document.createProcessingInstruction(name)`
- `Document.createAttribute(name)`
- `Document.createAttributeNS(namespaceURI, qualifiedName)`
- `Document.getElementById(tagName)`
- `Document.getElementByIdNS(namespaceURI, localName)`

## OBJETOS ELEMENTO

- `Element.tagName`
- `Element.getElementsByTagName(tagName)`
- `Element.getElementsByTagNameNS(namespaceURI, localName)`
- `Element.hasAttribute(name)`
- `Element.hasAttributeNS(namespaceURI, localName)`
- `Element.getAttribute(name)`
- `Element.getAttributeNS(namespaceURI, localName)`
- `Element.getAttributeNodeNS(namespaceURI, localName)`
- `Element.removeAttribute(name)`
- `Element.removeAttributeNode(oldAttr)`



## OBJETOS ATRIBUTO

- Attr.name
- Attr.localName
- Attr.prefix
- Attr.value

## OBJETOS NAMEDNODEMAP

- NamedNodeMap.length
- NamedNodeMap.(index)

## OBJETOS COMENTARIO

- Comment.data

## OBJETOS TEXTO Y CDATASECTION

- Text.data

## OBJETOS PROCESSINGINSTRUCTION

- ProcessingInstruccions.target
- ProcessingInstruction.data

## XPATH XML

### ¿QUE ES?

xPath es un módulo de Python que es parte de la librería `xml.etree.ElementTree`, este modulo 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 sí mismo.

### SINTAXIS

| Sintaxis                 | descripción   |
|--------------------------|---|
| <b>Tag</b>               | Selecciona todos los elementos contenidos en la etiqueta "tag"  |
| <b>*</b>                 | Selecciona todos los elementos hijos que posean este símbolo  |
| <b>.</b>                 | Selecciona el nodo actual, este es muy usado en el inicio del path, para indicar que es un path relativo.   |
| <b>//</b>                | Selecciona todos los subelementos de todos los niveles bajo el nodo expresado   |
| <b>..</b>                | Selecciona el elemento padre.   |
| <b>[@attrib]</b>         | Selecciona todos los elementos que contienen el atributo tras el "@"  |
| <b>[@attrib="value"]</b> | Seleccione todos los elementos para los cuales el atributo dado tenga un valor dado, el valor no puede contener comillas  |
| <b>[tag]</b>             | Seleccione todos los elementos que contienen una etiqueta hijo llamada tag. Solo los hijos inmediatos son admitidos.  |
| <b>[tag="text"]</b>      | Selecciona todos los elementos que tienen una etiqueta hijo llamada tag incluyendo descendientes que sean igual al texto dado.  |
| <b>[position]</b>        | 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 última o la posición relativa con respecto a la última posición last()-1 |