

UNIVERSIDAD SAN CARLOS DE GUATEMALA
INGENIERIA EN CIENCIAS Y SISTEMAS
IPC 2

ACTIVIDAD 2

PAULA GABRIELA GARCIA REINOSO

201700823

Contenido

DOM xml para Python	3
Descripción.....	3
Ejemplos	3
xPath XML para python	5
Descripción.....	5
Ejemplos	5
e-grafía.....	7

DOM xml para Python

Descripción

XML significa Extensible Markup Language (e X extensible M arkup L idioma). Usted puede aprender a través de este sitio Tutorial XML

XML está diseñado para transmitir y almacenar datos.

XML es un conjunto de reglas para definir la semántica de etiquetas, estas etiquetas documentará divididos en muchas partes y estas partes que ser identificados.

También es un lenguaje de metamarcado que define la sintaxis del lenguaje utilizado para definir otra, semántica, lenguaje de marcado con estructura de dominio específico.

Ejemplos

```
1. <?xml version="1.0"?>
2. <empresa>
3.   <empleado id="1">
4.     <nombre>José Ernesto</nombre>
5.     <username>jose</username>
6.     <password>321423</password>
7.   </empleado>
8.   <empleado id="2">
9.     <nombre>Daniel Pérez</nombre>
10.    <username>dperez</username>
11.    <password>433543</password>
12.  </empleado>
13. </empresa>
```

```

1.  from xml.dom import minidom
2.
3.  doc = minidom.parse("/ruta/datos.xml")
4.
5.  nombre = doc.getElementsByTagName("nombre")[0]
6.  print(nombre.firstChild.data)
7.
8.  empleados = doc.getElementsByTagName("empleado")
9.  for empleado in empleados:
10.     sid = empleado.getAttribute("id")
11.     username = empleado.getElementsByTagName("username")[0]
12.     password = empleado.getElementsByTagName("password")[0]
13.     print("id:%s " % sid)
14.     print("username:%s" % username.firstChild.data)
15.     print("password:%s" % password.firstChild.data)

```

```

from xml.dom.minidom import parse, parseString

dom1 = parse('c:\\temp\\mydata.xml') # parse an XML file by name

datasource = open('c:\\temp\\mydata.xml')
dom2 = parse(datasource) # parse an open file

dom3 = parseString('<myxml>Some data<empty/> some more data</myxml>')

```

```

from xml.dom.minidom import getDOMImplementation

impl = getDOMImplementation()

newdoc = impl.createDocument(None, "some_tag", None)
top_element = newdoc.documentElement
text = newdoc.createTextNode('Some textual content.')
top_element.appendChild(text)

```

```

1 <?xml version="1.0"?>
2 <Edit_Mensaje>
3 ..<Mensaje>
4 ....<Remitente>
5 .....<Nombre>Nombre del remitente</Nombre>
6 .....<Mail>Correo del remitente</Mail>
7 ....</Remitente>
8 ....<Destinatario>
9 .....<Nombre>Nombre del destinatario</Nombre>
10 .....<Mail>Correo del destinatario</Mail>
11 ....</Destinatario>
12 ....<Texto>
13 .....<Asunto>Este es mi documento con una estructura muy sencilla no contiene
    atributos ni entidades...</Asunto>
14 .....<Parrafo>Este es mi documento con una estructura muy sencilla no contiene
    atributos ni entidades...</Parrafo>
15 ....</Texto>
16 ..</Mensaje>
17 </Edit_Mensaje>
18 |

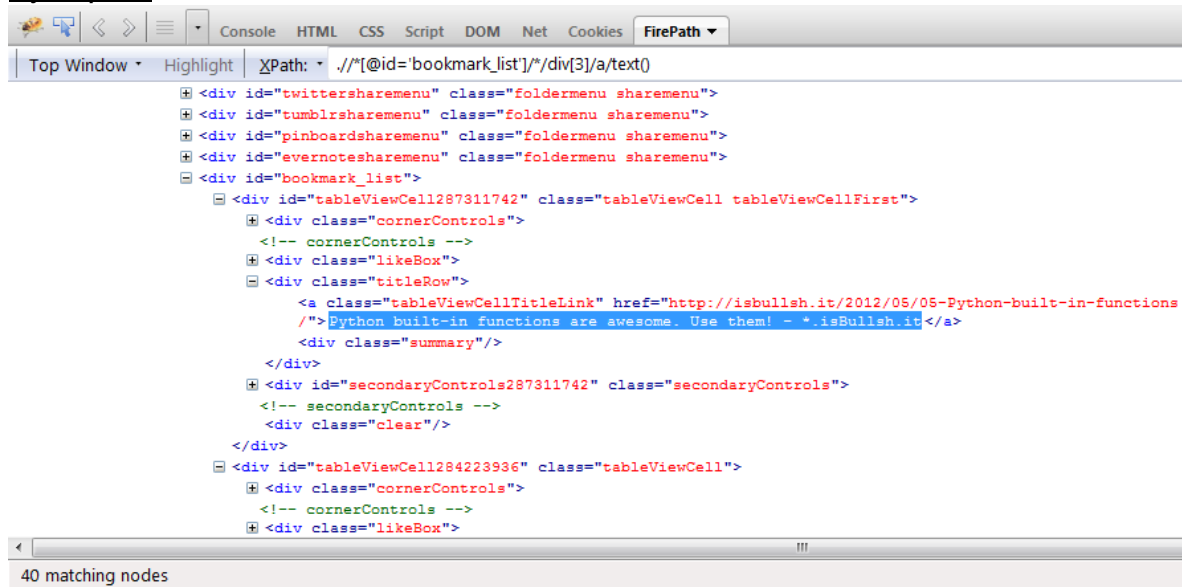
```

xPath XML para python

Descripción

XPath (XML Path Language) es un lenguaje que permite recuperar información de un documento XML. Para ello define una sintaxis para establecer partes en un documento XML, permitiendo navegar a través de sus elementos y atributos, además permite manipular de forma básica booleanos, números y cadenas.

Ejemplos



```

▼<div class="cardDetailsWrap">
  ▶<div class="detailImage">...</div>
  ▼<div class="cardDetails">
    <h1>Nidoqueen (7)</h1>
    ▶<div id="productPriceSummary">...</div>
    ▶<div id="alertOutOfStock" class="catalogAlert red" style="display: none;">...</div>
    ▶<div id="alertNoPricesBasedOnFilters" class="catalogAlert red priceTable" style="display: none;">...</div>
    <hr>
    ▼<table>
      ▼<tbody>
        ▼<tr>
          ▼<td>
            <b>Set Name:</b>
          </td>
          ▼<td>
            <a href="/pokemon/jungle">Jungle</a> == $0
          </td>
        </tr>
      </tbody>
    </table>
  </div>
  ▼<table border="0" cellspacing="0" cellpadding="0">
    ▼<tbody>
      ▶<tr align="left" class="header">...</tr>
      ▼<tr class="available" id="row185558-50G">
        ▼<td class="sku">
          <input id="Desc185558-50G" name="Description" type="hidden" value="BENZYL CHLORIDE, REAGENTPLUS(R), 99%, &">
          <input id="Fees185558-50G" name="Fees" type="hidden" value="Packaging Fee-Small Barrier Pack">
          <input id="CAS185558-50G" name="CasNumber" type="hidden" value="0000100447">
          <input id="Trans185558-50G" name="TransportationNumber" type="hidden" value="1738">
          <input id="JFC185558-50G" name="JapanFireCode" type="hidden" value="4.2">
          <input id="PSC185558-50G" name="PoisonSubstanceCode" type="hidden" value>
          <input id="NDC185558-50G" name="NDCNumber" type="hidden" value>
          <input id="Drug185558-50G" name="DrugType" type="hidden" value>
          <input id="Inquire185558-50G" name="Inquire" type="hidden" value>
          <input id="Tariff185558-50G" name="TariffCode" type="hidden" value="2903.99.2000">
          <p>185558-50G</p>
        </td>
        ▶<td class="shipping">...</td>
        <td align="center" class="safety">&nbsp;</td>
        ▼<td class="price" align="center">
          <p>26.50</p>
        </td>
      </tr>
    </tbody>
  </table>

```

```

def main():
    tree = etree.parse('feed.xml')
    NSMAP = {"nn": "http://www.w3.org/2005/Atom"}
    test = tree.xpath('//nn:category[@term="html"]/..', namespaces=NSMAP)
    for elem in tree.iter():
        print(elem.tag, '\t', elem.attrib)
    print('-----')
    test1 = tree.xpath('//nn:category', namespaces=NSMAP)
    print('+++++')
    for node in test1:
        test2 = node.xpath('../nn:summary', namespaces=NSMAP)
        print(test2[0].xpath('./text()[normalize-space()]'))
    print('*****')
    test3 = tree.xpath('//text()[normalize-space()]')# [normalize-space()] only remove the heading
    print(test3)

```

```

▼<div class="x-layer x-combo-list " id="ext-gen556" style="position:
absolute; z-index: 9010; visibility: visible; left: 104px; top: 321px; width:
204px; height: 300px; font-size: 12px;">
  ▼<div class="x-combo-list-inner" id="ext-gen557" style="width: 204px;
height: 300px;">
    <div class="x-combo-list-item">Heading</div>
  ... <div class="x-combo-list-item">Global ID</div> == $0
    <div class="x-combo-list-item">API ID</div>
    <div class="x-combo-list-item">Item Type</div>
    <div class="x-combo-list-item">Locked</div>
    <div class="x-combo-list-item">ID</div>

```

e-grafía

<http://www.w3big.com/es/python/python-xml.html>

<https://decodigo.com/python-leer-archivo-xml>

<https://docs.python.org/es/3.10/library/xml.dom.minidom.html>

[https://www.adictosaltrabajo.com/2009/02/26/teoria-xpath/#:~:text=XPath%20\(XML%20Path%20Language\)%20es,informaci%C3%B3n%20de%20un%20documento%20XML.&text=elementos%20y%20atributos%2C%20adem%C3%A1s%20permite,b%C3%A1sica%20booleanos%2C%20n%C3%BAmeros%20y%20cadena](https://www.adictosaltrabajo.com/2009/02/26/teoria-xpath/#:~:text=XPath%20(XML%20Path%20Language)%20es,informaci%C3%B3n%20de%20un%20documento%20XML.&text=elementos%20y%20atributos%2C%20adem%C3%A1s%20permite,b%C3%A1sica%20booleanos%2C%20n%C3%BAmeros%20y%20cadena)
s.