

Python

# Archivos XML



Rogelio Ferreira Escutia

Profesor / Investigador  
Tecnológico Nacional de México  
Campus Morelia



# Tipos de Archivos

# Tipos de Archivos

- **Básicamente hay 2 tipos de archivos:**

## Text File

- **Plain Text:** .txt, .csv
- **Source Code:** .py, .html, .css, .js
- **Data:** .json, .xml

## Binary File

- **Executable:** .exe, .dmg, .bin
- **Images:** .jpg, .png, .gif, .tiff, .ico
- **Video:** .mp4, .m4v, .mp4, .mov
- **Audio:** .aif, .mp3, .mpa, wav
- **Compressed:** .zip, .deb, .tar.gz
- **Font:** .woff, .otf, .ttf
- **Document:** .pdf, .docx, .xlsx

# XML

- **“XML”, siglas en inglés de “eXtensible Markup Language”, traducido como “Lenguaje de Mercado Extensible” o “Lenguaje de Marcas Extensible”.**
- **Es un metalenguaje que permite definir lenguajes de marcas desarrollado por el “World Wide Web Consortium” (W3C) utilizado para almacenar datos en forma legible.**

# XML

- **“XML”, siglas en inglés de “eXtensible Markup Language”, traducido como “Lenguaje de Mercado Extensible” o “Lenguaje de Marcas Extensible”.**
- **Es un metalenguaje que permite definir lenguajes de marcas desarrollado por el “World Wide Web Consortium” (W3C) utilizado para almacenar datos en forma legible.**

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE Edit_Mensaje SYSTEM "Edit_Mensaje.dtd">

<Edit_Mensaje>
  <Mensaje>
    <Remitente>
      <Nombre>Nombre del remitente</Nombre>
      <Mail> Correo del remitente </Mail>
    </Remitente>
    <Destinatario>
      <Nombre>Nombre del destinatario</Nombre>
      <Mail>Correo del destinatario</Mail>
    </Destinatario>
    <Texto>
      <Asunto>
        Este es mi documento con una estructura muy sencilla
        no contiene atributos ni entidades...
      </Asunto>
      <Parrafo>
        Este es mi documento con una estructura muy sencilla
        no contiene atributos ni entidades...
      </Parrafo>
    </Texto>
  </Mensaje>
</Edit_Mensaje>
```

# Ejemplo de una Factura del SAT en XML

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <cfdi:Comprobante xmlns:cfdi="http://www.sat.gob.mx/cfd/3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.sat.gob.mx/cfd/3 http://www.sat.gob.mx/sitio_internet/cfd/3/cfdv33.xsd" LugarExpedicion="58120"
  MetodoPago="PUE" TipoDeComprobante="I" Moneda="MXN" Total="60" SubTotal="60" NoCertificado="00001000000505491234"
  Fecha="2022-02-11T20:07:57" Folio="66123" Serie="A" FormaPago="01" Version="3.3" Sello="123/456/789=="
  Certificado="qwertyuiop1234567890">
3
4   <cfdi:Emisor RegimenFiscal="603" Nombre="INSTITUTO TECNOLOGICO DE MORELIA" Rfc="TNM140723GFA" />
5   <cfdi:Receptor Nombre="FERREIRA ESCUTIA ROGELIO" Rfc="SEP210905778" UsoCFDI="G03" />
6   <cfdi:Conceptos>
7
8     <cfdi:Concepto Importe="60.00" ValorUnitario="60.0000" Descripcion="Constancia de Inscripción" ClaveUnidad="A14"
9     Cantidad="1" ClaveProdServ="01010101">
10
11       <cfdi:Impuestos>
12         <cfdi:Traslados>
13           <cfdi:Traslado TipoFactor="Exento" Impuesto="002" Base="60.00" />
14         </cfdi:Traslados>
15       </cfdi:Impuestos>
16
17     </cfdi:Concepto>
18
19   </cfdi:Conceptos>
20
21
22   <cfdi:Complemento><tfd:TimbreFiscalDigital xmlns:tfd="http://www.sat.gob.mx/TimbreFiscalDigital"
  xsi:schemaLocation="http://www.sat.gob.mx/TimbreFiscalDigital http://www.sat.gob.mx/sitio_internet/cfd/
  TimbreFiscalDigital/TimbreFiscalDigitalv11.xsd" Version="1.1" UUID="5A37C367-FB6D-43DD-8553-121234567890"
  FechaTimbrado="2022-02-11T20:07:59" RfcProvCertif="BBA94070M2023" SelloCFD="abcd/efgh/ijkl=="
  NoCertificadoSAT="00001000001234567890" SelloSAT="1234567890asdfghijkl==" /></cfdi:Complemento></cfdi:Comprobante>

```

# Leer XML (1):

- **Bibliotecas utilizadas:**

```
# Bibliotecas utilizadas  
from xml.dom import minidom
```

# Leer XML (2):

- **Especificar archivo:**

```
# Especificar el archivo XML a utilizar  
xml= 'factura.xml'
```

# Leer XML (3):

- Parsear:

```
# Parsear el XML y generar un DOM  
mytree = minidom.parse(xml)
```

# Leer XML (4):

- Leer Emisor:

```
# Extrayendo datos del Emisor
etiquetas = mytree.getElementsByTagName('cfdi:Emisor')
for indice in etiquetas:
    emisor_nombre = indice.attributes['Nombre'].value
    print("\nNombre del Emisor: ", emisor_nombre)
    emisor_rfc = indice.attributes['Rfc'].value
    print("RFC del Emisor: ", emisor_rfc)
```

# Leer XML (5):

- **Leer Receptor:**

```
# Extrayendo datos del Receptor
etiquetas = mytree.getElementsByTagName('cfdi:Receptor')
for indice in etiquetas:
    receptor_nombre = indice.attributes['Nombre'].value
    print("\nNombre del Receptor: ", receptor_nombre)
    receptor_rfc = indice.attributes['Rfc'].value
    print("RFC del Receptor: ", receptor_rfc)
```

# Leer XML (6):

- Leer datos de la factura:

```
# Extrayendo datos de la Factura
etiquetas = mytree.getElementsByTagName('cfdi:Concepto')
for indice in etiquetas:
    importe = indice.attributes['Importe'].value
    print("\nImporte: ", importe)
    descripcion = indice.attributes['Descripcion'].value
    print("Descripción: ", descripcion, "\n")
```

# Leer XML (7):

- **Salida:**

Nombre del Emisor: INSTITUTO TECNOLOGICO DE MORELIA  
RFC del Emisor: TNM140723GFA

Nombre del Receptor: FERREIRA ESCUTIA ROGELIO  
RFC del Receptor: SEP210905778

Importe: 60.00  
Descripción: Constancia de Inscripción



## Rogelio Ferreira Escutia

Profesor / Investigador  
Tecnológico Nacional de México  
Campus Morelia



[rogelio.fe@morelia.tecnm.mx](mailto:rogelio.fe@morelia.tecnm.mx)



[rogeplus@gmail.com](mailto:rogeplus@gmail.com)



[xumarhu.net](http://xumarhu.net)



[@rogeplus](https://twitter.com/rogeplus)



[https://www.youtube.com/  
channel/UC0on88n3LwTKxJb8T09sGjg](https://www.youtube.com/channel/UC0on88n3LwTKxJb8T09sGjg)



[rogelioferreiraescutia](https://www.linkedin.com/in/rogelioferreiraescutia)

