

Formulas

March 12, 2021

Aceso al xml con python - lxml etree

Iteración

```
[1]: from lxml import etree
from io import StringIO

xml_text="<?xml version='1.0'?>\
    <person id='Técnico'>\
        <name>Diego Coppa</name>\
        <eMail>dcoppa@inti.gob.ar</eMail>\
        <phone>(5411) 4724-6200 int 6933</phone>\
        <location>\
            <city>Provincia de Buenos Aires</city>\
            <countryCode>AR</countryCode>\
            <postCode>B1650WAB</postCode>\
            <state>Argentina</state>\
            <street>Av. General Paz</street>\
            <streetNo>5445</streetNo>\
            <further>\
                <content>San Martín</content>\
            </further>\
        </location>\
    </person> "
```

```
def parseXML():
    elem_tree=etree.parse(StringIO(xml_text))
    for tag in elem_tree.iter():
        if not len(tag):
            print(tag.tag , "-> ", tag.text)

if __name__ == "__main__":
    parseXML()
```

```
name -> Diego Coppa
eMail -> dcoppa@inti.gob.ar
phone -> (5411) 4724-6200 int 6933
```

```
city -> Provincia de Buenos Aires
countryCode -> AR
postCode -> B1650WAB
state -> Argentina
street -> Av. General Paz
streetNo -> 5445
content -> San Martín
```

Por path

```
[9]: from lxml import etree
from io import StringIO
import dateutil.parser

xml_text="<?xml version='1.0'?>\
    <person id='Técnico'?>\
        <name>Diego Coppa</name>\
        <eMail>dcoppa@inti.gob.ar</eMail>\
        <phone>(5411) 4724-6200 int 6933</phone>\
        <location>\
            <city>Provincia de Buenos Aires</city>\
            <countryCode>AR</countryCode>\
            <postCode>B1650WAB</postCode>\
            <state>Argentina</state>\
            <street>Av. General Paz</street>\
            <streetNo>5445</streetNo>\
            <further>\
                <content>San Martín</content>\
            </further>\
        </location>\
    </person> "
```

```
path="/person/phone"

def parseXML():
    elem_tree=etree.parse(StringIO(xml_text))
    r=elem_tree.xpath(path)
    print(r[0].tag, "-> ", r[0].text)

if __name__ == "__main__":
    parseXML()
```

```
phone -> (5411) 4724-6200 int 6933
```

Archivo xml

```
[8]: from lxml import etree
from io import StringIO, BytesIO
```

```

import dateutil.parser

#path_a="/dcc:digitalCalibrationCertificate/dcc:administrativeData/dcc:
    ↳calibrationLaboratory/dcc:contact[@id='INTI']/dcc:name/dcc:
    ↳content[@lang='es']"
#fecha de recibimiento del equipo
#path_a="/dcc:digitalCalibrationCertificate/dcc:administrativeData/dcc:coreData/
    ↳dcc:receiptDate"
#magnitud
#path_a="/dcc:digitalCalibrationCertificate/dcc:measurementsResults/dcc:list/
    ↳dcc:quantity/si:real/si:value"
#magnitud
path_a="/dcc:digitalCalibrationCertificate/dcc:measurementsResults/dcc:list/dcc:
    ↳quantity/si:real/si:unit"
#incertidumbre de medición
#path_a="/dcc:digitalCalibrationCertificate/dcc:measurementsResults/dcc:list/
    ↳dcc:quantity/si:real/si:expandedUnc/si:uncertainty"
#path al laboratorio o el nmi que creó el certificado

def parseXML(xmlFile):
    with open(xmlFile) as fobj:
        xml_str = fobj.read() # devuelve string
        elem_tree=etree.parse(StringIO(xml_str)) # devuelve elementr tree
        r=elem_tree.xpath(path_a, namespaces={'dcc': 'https://ptb.de/dcc', 'si':↳
    ↳'https://ptb.de/si' })
        data=str(r[0].tag)
        print(data[data.find("{}")+1:], "-> ", r[0].text)

        #print(r)

if __name__ == "__main__":
    parseXML("xml/calibracion2.xml")

```

unit -> \kilogram

[]: