

Artículo

[Ricardo Paiva](#) · 18 ene, 2023 · Lectura de 4 min

Validación de documentos XML con Schematron mediante Python

Schematron es un lenguaje de validación basado en reglas para hacer aserciones/afirmaciones sobre la presencia o ausencia de ciertos patrones en documentos XML. Un Schematron se refiere a una colección de una o más reglas que contienen pruebas. Los Schematron están escritos en una forma de XML, lo que los hace relativamente fáciles de inspeccionar, comprender y escribir para todos, incluso los que no son programadores.

Esencialmente, un Schematron realiza dos acciones en secuencia:

Encuentra nodos de contexto de interés en el documento. Un "nodo de contexto" puede ser un elemento de un tipo particular o un elemento específico en un lugar determinado del documento, un atributo o un valor de atributo. Por ejemplo, supone que quieres verificar si la suma de los elementos <Percent> dentro de cada un nodo contexto es de 100%. En este caso, el nodo de contexto sería el elemento <Total>. Para cada uno de esos nodos de interés, comprueba si una declaración específica es verdadera o falsa. Por ejemplo, puede tener una regla escrita para responder a la pregunta "¿Es la suma total de 100%?".

El recurso idóneo para buscar más detalle sobre el tema sería: <https://www.schematron.com/>. Para nosotros lo que importa es que podamos validar nuestro documento XML en base a una definición Schematron. Para ello hay que tener en cuenta que hay múltiples proyectos open source con implementaciones de Schematron para XSLT. Uno de los más interesantes lo tenéis disponible en <https://github.com/schxslt/schxslt.git>.

En este artículo se pretende apalancar las capacidades de Python disponibles en InterSystems IRIS (for Health) o HealthShare (Health Connect).

Para ello necesitamos una instancia de InterSystems IRIS o HealthShare Health Connect. Para nuestro ejemplo usaremos un contenedor con la última community edition de InterSystems IRIS for Health. Hemos de arrancar la instancia publicando las puertas default y mapeando el directorio corriente a la carpeta durable en el container.

```
$ docker run --name iris4health -d --publish 51773:51773 --publish 52773:52773 --volume $(pwd):/durable  
containers.intersystems.com/intersystems/irishealth-community:2022.3.0.589.0
```

Ahora que ya tenemos nuestra instancia en ejecución podremos arrancar una consola en el container.

```
$ docker exec -it <containerID> bash
```

Ahora ya nos podemos dedicar al modulo Python. Usaremos lxml. Se trata de un enlace Python para las bibliotecas C libxml2 y libxslt. Es único en el sentido de que combina la velocidad y la integridad de las funciones XML de estas bibliotecas con la simplicidad de una API Python nativa, en su mayoría compatible con la conocida API de ElementTree. Para más información sobre lxml <https://lxml.de/index.html>

Asumiendo que el gestor de paquetes pip3 (y por supuesto Python 3) ya está instalado en la instancia, habrá que instalar el modulo debido.

```
$ pip3 install --target /usr/irissys/mgr/python lxml
```

El método ejemplo que usaremos estará codificado en Python y estará encargado del parseo y validación de las reglas del schematron. El código de la clase que usaremos es el siguiente:

```
Class dc.schematron Extends %RegisteredObject
{

/// Description
ClassMethod simpleTest() [ Language = python ]
{
    from lxml import isoschematron
    from lxml import etree

    print("Validating File...\n")

    # def runsch(rulesFile, xmlFile):
    #open files
    rules = open('/durable/test-schema.sch', 'rb') # Schematron schema
    XMLhere = open('/durable/test-file.xml', 'rb') # XML file to check
    #Parse schema
    sct_doc= etree.parse(rules)
    schematron=isoschematron.Schematron(sct_doc, store_report=True)

    #Parse XML
    doc = etree.parse(XMLhere)

    #Validate against schema
    validationResult = schematron.validate(doc)
    report = schematron._validation_report

    #Check result
    if validationResult:
        print("passed")
    else:
        print("failed")
        print(report)
}
}
```

La verdad es que se trata de un método bastante sencillo. Abre 2 ficheros – el fichero con las reglas (schematron) y un fichero ejemplo. La regla es verificar si la suma de los elementos <Percent> dentro de cada nodo <Total> es de 100%. Para ejecutarlo habrá que lanzar el siguiente comando desde la consola:

```
d ##class(dc.schematron).simpleTest()
```

El resultado se presentará en la consola. La misma lógica se podrá usar en una producción de interoperabilidad.

El código fuente conteniendo todos los elementos esta disponible [aquí](#).

[#Python](#) [#XML](#) [#HealthShare](#) [#InterSystems IRIS](#) [#InterSystems IRIS for Health](#)

URL de
fuente: <https://es.community.intersystems.com/post/validaci%C3%B3n-de-documentos-xml-con-schematron-mediante-python>
