Deep Learning mit MATLAB

E-Book herunterladen und die Möglichkeiten von
Artificial Intelligence entdecken.

MathWorks

Home	Forums	Reviews	Tutorials	Articles	Register

LinuxQuestions.org > Forums > Non-*NIX Forums > Programming

[SOLVED] XML/XSD Schemavalidation of an OOXML document

User Name User Name
Password

Programming This forum is for all programming questions.
The question does not have to be directly related to Linux and any language is fair game.

Notices

Welcome to **LinuxQuestions.org**, a friendly and active Linux Community.

You are currently viewing LQ as a guest. By joining our community you will have the ability to post topics, receive our newsletter, use the advanced search, su
and access many other special features. Registration is quick, simple and absolutely free. **Join our community** today!

Note that registered members see fewer ads, and ContentLink is completely disabled once you log in.

Are you new to LinuxQuestions.org? Visit the following links:
Site Howto | Site FAQ | Sitemap | Register Now

If you have any problems with the registration process or your account login, please contact us. If you need to reset your password, click here.

Having a problem logging in? Please visit this page to clear all LQ-related cookies.

Introduction to Linux - A Hands on Guide

This guide was created as an overview of the Linux Operating System, geared toward new users as an exploration tour and getting started
exercises at the end of each chapter. For more advanced trainees it can be a desktop reference, and a collection of the base knowledge nee
with system and network administration. This book contains many real life examples derived from the author's experience as a Linux system
administrator, trainer and consultant. They hope these examples will help you to get a better understanding of the Linux system and that yo
to try out things on your own.

Click Here to receive this Complete Guide absolutely free.

CDN Prices Revealed

Post Reply

03-16-2019, 09:06 AM

Michael Uplawski
Member

Registered: Dec 2015
Location: Outside, most of the time.
Posts: 830
Blog Entries: 25

Rep: ▪▪▪▪▪

XML/XSD Schemavalidation of an OOXML document

Good afternoon.

I am generating programmatically OOXML-documents for routine-use. As my knowledge of OOXML bases entirely on online-resource
and would like to validate the code in my template files (document.xml, header.xml) and styles (styles.xml) against the referenced
definitions, which are for now only these three:

- xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main"
- xmlns:mc="http://schemas.openxmlformats.org/markup-compatibility/2006"
- xmlns:wpc="http://schemas.microsoft.com/office/word/2010/wordprocessingCanvas"

My knowledge of **xmllint** is insufficient and online-validators appear to validate each of my documents as valid, even where I close
before one of the elements that it must include. The only thing that they achieve is assure the "*well-formedness*" of the XML.

Can you point me at a resource which explains how this type of document is best validated against the named schemas? Or where
the xsd for each schema, if I want to feed them to xmllint?

Amongst others, I have seen:

1. http://www.datypic.com/sc/ooxml/ss.html gives the same kind of overview that I find elsewhere, but no XSD-code.
2. https://www.ecma-international.org/p...s/Ecma-376.htm - I do not know what to download, here, an attempt to validate against "*ECMA-376 5th edition Part 1*" fails with the following errors:

Code:

```
user@machine:~$ xmllint --schema wml.xsd styles.xml
shared-math.xsd:154: element attribute: Schemas parser error : attribute use (unknown), attribute 're
wml.xsd:1663: element attribute: Schemas parser error : attribute use (unknown), attribute 'ref': The
WXS schema wml.xsd failed to compile

(...)
```

Last edited by Michael Uplawski; 03-16-2019 at 09:34 AM. Reason: three schemas, not four.

03-17-2019, 12:37 PM

NevemTeve
Senior Member

Registered: Oct 2011
Location: Budapest
Distribution:
Debian/GNU/Linux, AIX
Posts: 3,979

Rep: ⬛⬛⬛⬛⬛⬛⬛⬛

Could you please give an example xml (or a link to it)?

03-17-2019, 02:10 PM

Michael Uplawski
Member

Registered: Dec 2015
Location: Outside, most of the time.
Posts: 830

Original Poster
Blog Entries: 25

Rep: ⬛⬛⬛⬛⬛

Quote:

> Originally Posted by **NevemTeve** ▷
> *Could you please give an example xml (or a link to it)?*

Here is an archive with authentic "*templates*", i.e. a document.xml which is completed during the execution of my routine, a styles and a _rels directory, which links header.xml and document.xml.

I do not know if that is the kind of example you wish to see.

03-18-2019, 01:22 AM

NevemTeve
Senior Member

Registered: Oct 2011
Location: Budapest
Distribution:
Debian/GNU/Linux, AIX
Posts: 3,979

Rep: ⬛⬛⬛⬛⬛⬛⬛⬛

Ok, let's try document.xml

Code:

```
<?xml version="2.0" encoding="utf-8" standalone="yes"?>
<w:document xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main">
  <w:body>
    <w:type w:val="continuous" />
    <w:sectPr>
      <w:headerReference w:type="default" r:id="rId1" />
...
```

Well, here is the first problem reported by xmllint:

Code:

```
document.xml:1: parser error : Unsupported version '2.0'
<?xml version="2.0" encoding="utf-8" standalone="yes"?>
```

the second one:

Code:

```
document.xml:6: namespace error : Namespace prefix r for id on headerReference is not defined
        <w:headerReference w:type="default" r:id="rId1" />
```

The trivial fix would be this:

Code:

```
<?xml version="1.0" encoding="utf-8" standalone="yes"?>
<w:document
    xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main"
    xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships">
  <w:body>
...
```

But that's not enough: as xmllint accepts only one schema, one should create an xml-catalog. I'll try to give more details.

PS: I found the xsd files here: https://jar-download.com/cache_jars/.../jar_files.zip

Last edited by NevemTeve; 03-18-2019 at 02:19 AM.

1 members four

Michael Uplawski
Member

Registered: Dec 2015
Location: Outside, most of the time.
Posts: 830

Original Poster
Blog Entries: 25

Rep: 🟩🟩🟩🟩🟩

Quote:
> Originally Posted by **NevemTeve** ▶
> *But that's not enough: as xmllint accepts only one schema, one should create an xml-catalog. I'll try to give more details.*
>
> *PS: I found the xsd files here: https://jar-download.com/cache_jars/.../jar_files.zip*

Thank you for your time and effort. I was impatient to read your response.

Are you accustomed to this kind of problem or how did you make a connection to jar-download.com? Even if XML and Java are clos hope for a generally applicable procedure and would not have thought of searching for a jar-archive, of all choices... called zip, if it

Last edited by Michael Uplawski; 03-18-2019 at 07:11 AM.

NevemTeve
Senior Member

Registered: Oct 2011
Location: Budapest
Distribution: Debian/GNU/Linux, AIX
Posts: 3,979

Rep: 🟩🟩🟩🟩🟩🟩🟩🟩

Well, I'd suggest this:

As root
1. If you don't have file /usr/local/etc/xml/catalog, create it:

Code:
```
$ mkdir -p /usr/local/etc/xml
$ cat >/usr/local/etc/xml/catalog <<DONE
<?xml version="1.0"?>
<!DOCTYPE catalog PUBLIC "-//OASIS//DTD Entity Resolution XML Catalog V1.0//EN" "http://www.oasis-open.org/
<catalog xmlns="urn:oasis:names:tc:entity:xmlns:xml:catalog">
  <nextCatalog catalog="file:///etc/xml/catalog"/>
</catalog>
DONE
```

2. If you don't have a line in this file referring to name="http://www.w3.org/XML/1998/namespace", then insert line like this (befc line):

Code:
```
    <uri name="http://www.w3.org/XML/1998/namespace" uri="file:///usr/local/etc/xml/xml_2009_01.xsd"/>
```

Also actually download this file:

Code:
```
wget -O /usr/local/etc/xml/xml_2009_01.xsd http://www.w3.org/2009/01/xml.xsd
```

Switch back to normal user.
3. Put the OOXML-xsd files into a sub-directory of your work-dir, eg ooxml_xsd.

4. Some modifications are required to let xmllint work:
4.1. wml.xsd -- missing schemaLocation

Code:
```
-   <xsd:import id="xml" namespace="http://www.w3.org/XML/1998/namespace" />
+   <xsd:import id="xml" namespace="http://www.w3.org/XML/1998/namespace" schemaLocation="http://www.w3.org/
```

4.2. dml-wordprocessingDrawing.xsd -- duplicate import for the same namespace

Code:
```
-   <xsd:import schemaLocation="dml-graphicalObject.xsd"    namespace="http://schemas.openxmlformats.org/dra
-   <xsd:import schemaLocation="dml-documentProperties.xsd" namespace="http://schemas.openxmlformats.org/dra
+   <xsd:import schemaLocation="dml-wordprocessingDrawing_import.xsd" namespace="http://schemas.openxmlforma
```

Then create this dml-wordprocessingDrawing_import.xsd file:

Code:
```
<?xml version="1.0" encoding="utf-8"?>
<xsd:schema targetNamespace="http://schemas.openxmlformats.org/drawingml/2006/main"
   elementFormDefault="qualified"
   xmlns:xsd="http://www.w3.org/2001/XMLSchema">
   <xsd:include schemaLocation="dml-graphicalObject.xsd"/>
   <xsd:include schemaLocation="dml-documentProperties.xsd"/>
</xsd:schema>
```

Now you can invoke xmllint:

Code:
```
$ export XML_CATALOG_FILES=/usr/local/etc/xml/catalog
$ xmllint -noout -debugent -schema ooxml_xsd/wml.xsd document.xml

new input from file: ooxml_xsd/wml.xsd
new input from file: ooxml_xsd/shared-customXmlSchemaProperties.xsd
new input from file: ooxml_xsd/shared-math.xsd
new input from file: ooxml_xsd/dml-wordprocessingDrawing.xsd
new input from file: ooxml_xsd/dml-wordprocessingDrawing_import.xsd
new input from file: ooxml_xsd/dml-graphicalObject.xsd
new input from file: ooxml_xsd/dml-documentProperties.xsd
new input from file: ooxml_xsd/dml-baseTypes.xsd
```

```
new input from file: ooxml_xsd/shared-relationshipReference.xsd
new input from file: ooxml_xsd/dml-shapeGeometry.xsd
new input from file: file:///usr/local/etc/xml/xml_2009_01.xsd
new input from file: document.xml
document.xml:6: element type: Schemas validity error :
Element '{http://schemas.openxmlformats.org/wordprocessingml/2006/main}type':
This element is not expected.
Expected is ( {http://schemas.openxmlformats.org/wordprocessingml/2006/main}sectPr ).
document.xml fails to validate
DOCUMENT
No entities in internal subset
No entities in external subset
```

Last edited by NevemTeve; 03-18-2019 at 08:04 AM.

2 members four

■ 03-20-2019, 05:34 PM

Michael Uplawski
Member

Registered: Dec 2015
Location: Outside, most of the time.
Posts: 830

Original Poster
Blog Entries: 25

Rep: 🟩🟩🟩🟩🟩

Thank you so much!

I validate.
And promptly, my "templates" must be revised as I have skipped some namespaces, at least for the attributes of the *<w:headerRe*
now I was lucky that the text-processor, which reads my final documents, corrects errors up on saving and my requirements were s

It is, however, surprising that the validation needs so much preparation.

Last edited by Michael Uplawski; 03-23-2019 at 04:20 AM. Reason: upon

1 members four

■ 07-28-2019, 06:30 PM

Ricky Rocker
LQ Newbie

Registered: Jul 2019
Posts: 3

Rep: ■

😕

Hi @nevemTeve

I "think" I've got everything right (I've modified your approach slightly by placing the xml_2009_01.xsd file in the same folder as t
testing hopefully without the requirement for the catalog)

and I'm getting the following...

Code:
```
root@dev:/Development # xmllint --schema /Development/OfficeOpenXML-XMLSchema-Strict/wml.xsd testdoc.xml --
new input from file: /Development/OfficeOpenXML-XMLSchema-Strict/wml.xsd
new input from file: /Development/OfficeOpenXML-XMLSchema-Strict/dml-wordprocessingDrawing.xsd
new input from file: /Development/OfficeOpenXML-XMLSchema-Strict/dml-main.xsd
new input from file: /Development/OfficeOpenXML-XMLSchema-Strict/shared-relationshipReference.xsd
new input from file: /Development/OfficeOpenXML-XMLSchema-Strict/shared-commonSimpleTypes.xsd
new input from file: /Development/OfficeOpenXML-XMLSchema-Strict/dml-diagram.xsd
new input from file: /Development/OfficeOpenXML-XMLSchema-Strict/dml-chart.xsd
new input from file: /Development/OfficeOpenXML-XMLSchema-Strict/dml-chartDrawing.xsd
new input from file: /Development/OfficeOpenXML-XMLSchema-Strict/dml-picture.xsd
new input from file: /Development/OfficeOpenXML-XMLSchema-Strict/dml-lockedCanvas.xsd
new input from file: /Development/OfficeOpenXML-XMLSchema-Strict/shared-math.xsd
new input from file: /Development/OfficeOpenXML-XMLSchema-Strict/xml_2009_01.xsd
new input from file: /Development/OfficeOpenXML-XMLSchema-Strict/shared-customXmlSchemaProperties.xsd
new input from file: testdoc.xml
testdoc.xml:1: element document: Schemas validity error : Element '{http://schemas.openxmlformats.org/wordp
testdoc.xml fails to validate
DOCUMENT
No entities in internal subset
No entities in external subset
root@dev:/Development #
```

...so the schemas all seem happy, but for some reason testdoc.xml with the following w:document node is failing as above (XMLsp

Code:
```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<w:document xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main"
xmlns:wpc="http://schemas.openxmlformats.org/office/word/2010/wordprocessingCanvas"
xmlns:mc="http://schemas.openxmlformats.org/markup-compatibility/2006"
xmlns:o="urn:schemas-microsoft-com:office:office"
xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships"
xmlns:m="http://schemas.openxmlformats.org/officeDocument/2006/math"
xmlns:v="urn:schemas-microsoft-com:vml"
xmlns:wp14="http://schemas.openxmlformats.org/office/word/2010/wordprocessingDrawing"
xmlns:wp="http://schemas.openxmlformats.org/drawingml/2006/wordprocessingDrawing"
xmlns:w10="urn:schemas-microsoft-com:office:word"
xmlns:w14="http://schemas.openxmlformats.org/office/word/2010/wordml"
xmlns:w15="http://schemas.openxmlformats.org/office/word/2012/wordml"
xmlns:wpg="http://schemas.openxmlformats.org/office/word/2010/wordprocessingGroup"
xmlns:wpi="http://schemas.openxmlformats.org/office/word/2010/wordprocessingInk"
xmlns:wne="http://schemas.openxmlformats.org/office/word/2006/wordml"
xmlns:wps="http://schemas.openxmlformats.org/office/word/2010/wordprocessingShape">
```

Any ideas would be greatly appreciated!

thanks so much

Ricky

Last edited by Ricky Rocker; 07-29-2019 at *04:36 PM.*

□ 07-29-2019, 03:56 PM

NevemTeve
Senior Member

Registered: Oct 2011
Location: Budapest
Distribution:
Debian/GNU/Linux, AIX
Posts: 3,979

Rep: ⬛⬛⬛⬛⬛⬛⬛⬛⬛

You might want to edit your post to add [code] and [/code] tags.

□ 07-29-2019, 04:40 PM

Ricky Rocker
LQ Newbie

Registered: Jul 2019
Posts: 3

Rep: ⬛

🙂

Hi,

Thanks very much for the formatting tip.

I've tidied up but no longer need assistance as have resolved the issue I explain above. It was actually a PHP DOMDocument issue PHP bug.

schemaValidate ignores namespaces dynamically added to a DOMDocument https://bugs.php.net/bug.php?id=78352

Last edited by Ricky Rocker; 07-29-2019 at *04:52 PM.*

1 members four

Post Reply

Tags

ooxml, validate, word-processor, xml, xsd

Posting Rules ⊗

You **may not** post new threads
You **may not** post replies
You **may not** post attachments
You **may not** edit your posts

BB code is **On**
Smilies are **On**
[IMG] code is **Off**
HTML code is **Off**

Forum Rules

Similar Threads

Similar Threads

Thread	Thread Starter	Forum	Replies	La
LXer: Rob Weir's OOXML Update, Part III - Making OOXML Conform to Office 2007	LXer	Syndicated Linux News	0	
LXer: South Africa Files Official Appeal Re OOXML - OOXML in Limbo Now	LXer	Syndicated Linux News	0	
Java XML client from xsd	true_atlantis	Programming	1	
trying to execute a .xsd file	kag291	Linux - Newbie	1	
java - verifying XSD	0raven0	Programming	1	

LinuxQuestions.org > Forums > Non-*NIX Forums > Programming

All times are GMT -5. The time now is 07:15 AM.

Contact Us - **Advertising Info** - **Rules** - **LQ Merchandise** - **Donations** - **Contributing M**
Open Source Consulting | Domain Registration