DocBook XML 4.1.2 Quick Start Guide

Jim Weller
Sleepless Tech

jim@sleeplesstech.com

Describes how to install, configure and use the tools and resources for DocBook XML 4.1.2. The purpose of this quick start guide is to get new docbook authors, editors, and contributors up and running fast with the DoocBook tools. These are powerful tool in the hands of an author. It assumes a fair knowledge of building and installing source packages. There are probably a million and one ways to accomplish my ultimate goal of installing and using these tools. This one works well for me. If you find this useful, please drop me an email, so I can brag to my Grand Parents.

Table of Contents

1. Legal Notice

2. Introduction

DocBook is a widely used DTD in SGML and XML that is tailored toward technical manuals. It's basically a way that authors can write a document once and then using SGML and XML tools to convert to many common formats (HTML, DOC, RTF, PDF, PostScript etc.). It is used on a huge number of open source projects as the main documentation system.

I'm working on a HOWTO/book. These are my notes from a three month bout with these tools and concepts. I had a difficult and frustrating time setting up these tools. Plus, I couldn't find a working equivalent. So, I truly hope this information will be helpful, but it is provided without warranty or gaurantee. If you break anything you get to keep both pieces.

This guide tends toward the DocBook spirit. I assume authors don't need to get mired in a gajillion layers of DocBook complexity just to generate an articulate piece of work. More simply: Get tools, write content and distribute.

Please, note that I do not cover the installation or usage of some of the 'fluffy' tools like docbook2X or sgml-tools. Nor do I make mention of backends other than html and text. I leave it as an exercise for the reader to generate different types of output (tex, pdf, etc.). This document is intended to be the straightest path to writing and evaluating your documents.

3. Requirements

You'll have to download and install a number of packages. Most won't take much to install, but you should be familiar with installing GNU source packages.

3.1. Basics

 You'll need a computer running linux with the GNU development environment :) You can do it with another OS, but you'll have to interpolate as necessary.

3.2. OpenJade

· OpenJade http://openjade.sourceforge.net

 You'll need to download a recent version of the openjade suite. OpenJade is an implementation of the ISO/IEC 10179:1996 standard DSSSL language.

3.3. DocBook DTDs

· SGML 3.1 http://www.oasis-open.org/docbook/sgml/3.1/index.html

· SGML 4.1 http://www.oasis-open.org/docbook/sgml/4.1/index.html

· XML 4.1.2 http://www.oasis-open.org/docbook/xml/4.1.2/index.html

 You'll need to get the zip archives from the sites listed above. DTD stands for Document Type Definition. From Oasis:

> DocBook is a DTD maintained by the DocBook Technical Committee of OASIS. It is particularly well suited to books and papers about computer hardware and software (though it is by no means limited to these applications).

3.4. ISO8879 Entities

· ISOEnts.zip http://www.oasis-open.org/cover/ISOEnts.zip

· isoENT-tar.gz http://www.oasis-open.org/cover/isoENT-tar.gz

· Missing Entity (iso-grk4.gml) iso-grk4.gml.gz

You'll need to get ISOEnts.zip and isoENT-tar.gz. These are all the symbol entities (e.g. ©,® and *lots* more). You can read about these at the SGML/XML Entity Sets and Entity Management (http://www.oasis-open.org/cover/topics.html#entities) site. If you understand all that, you probably don't need this guide ;-)

I kept getting an error message about a missing entity. It was no where to be found in any of my downloads. Greg Ferguson was nice enough to post if for me. I've mirrored it with this guide. You'll need to put it with the rest of your entities.

3.5. DocBook Style Sheets

· Norman Walsh's Modular DocBook Stylesheets http://nwalsh.com/docbook/dsssl/

· The LDP Customized Stylesheet http://www.linuxdoc.org/authors/tools/ldp.dsl

You'll need to download the nwalsh style sheets. You can get by with just that, but the LDP ones have some nice extensions. From Norman Walsh's docbook site:

> DSSSL is a stylesheet language for both print and online rendering. The acronym stands for Document Style Semantics and Specification Language. It is defined by ISO/IEC 10179:1996.

4. Installation and Configuration

Setting up the tools was the hardest part for me. The maze of catalogs, programs and acronyms is daunting to the first time (even technically minded) user. Never mind the miles of error message these tools generate, but I'll never cry RPM. I ran into a couple of snags with work arounds that I'll mention here.

4.1. Ground work

First setup some handy environmental variables.

```
export SGMLHOME=/usr/local/sgml
export DBARCHIVE=/where/you/put/your/source/downloads
mkdir $SGMLHOME
```

4.2. OpenJade

The only package that requires compiling is openjade. Openjade is a reasonable GNU autotools source build. Be sure to add openjade's lib directory to your library search path (ld.so.conf on Linux). Also, not that you have to manually copy the dsssl subdirectory as the install doesn't do it for you.

```
cd /usr/src
tar -xzf $DBARCHIVE/openjade-1.3.tar.gz
cd openjade-1.3
./configure --prefix=$SGMLHOME/openjade-1.3 && make && make install
cp -a dsssl $SGMLHOME/openjade-1.3
cd $SGMLHOME/openjade-1.3
mkdir lib
mv libo* lib
```

4.3. DocBook DTDs and Entities

Next, you'll need to unpack all the DocTools DTDs and entities. Some of the entities need to be renamed from .ent to .gml. You can rename by hand or try my crazy one line script in your shell. Once they are all renamed, you'll get them into the 3.1sgml and 4.1sgml directories respectively . Some of the tidy people in the world like their catalog files called 'catalog'. It can't be a bad convention. So I symlinked mine.

```
gunzip $DBARCHIVE/iso-grk4.gml.gz
cd $SGMLHOME
mkdir docbook
cd docbook
mkdir 3.1sgml 4.1sgml 4.1.2xml

cd 3.1sgml
unzip -a $DBARCHIVE/docbk31.zip
unzip -a $DBARCHIVE/ISOEnts.zip
cp $DBARCHIVE/iso-grk4.gml .
```

```
ln -s docbook.cat catalog

cd ../4.1sgml/
unzip -a $DBARCHIVE/docbk41.zip
unzip -a $DBARCHIVE/ISOEnts.zip
cp $DBARCHIVE/iso-grk4.gml .
ln -s docbook.cat catalog

cd ../4.1.2xml/
unzip -a $DBARCHIVE/docbkx412.zip
ln -s docbook.cat catalog

mkdir /tmp/ents
cd /tmp/ents
tar -xzf $DBARCHIVE/isoENT-tar.gz
for i in `find . -type f`;do mv $i `echo $i | sed -e 's/.ent/.gml/g'`;done
cp * $SGMLHOME/docbook/3.1sgml/
cp * $SGMLHOME/docbook/4.1sgml/
```

4.4. Norm Walsh's and LDP's Style Sheets

Now, we're ready to unpack Norm Walsh's and the LDP's Styles Sheets. These are wonderfully straight forward.

```
cd $SGMLHOME
mkdir dsssl
cd dsssl
unzip -a $DBARCHIVE/db164.zip
cd docbook/

cp $DBARCHIVE/ldp.dsl html/
cp $DBARCHIVE/ldp.dsl print/
```

4.5. Configuration

Then, well need to setup our SGML_CATALOG_FILES evironment variable. This is a list of files that openjade will use to find 'stuff'. I consolidated them all down to one file shown below. Once you get it set and working, I recommend making this a permanent fixture in your environment (.profile,.bashrc, etc.). Then when you want to change from 4.1.2 to 3.1 to compile other HOWTOS you just edit 'your' main catalog file.

```
export SGML_CATALOG_FILES=$SGMLHOME/catalog
```

$SGMLHOME/catalog (where /usr/local/sgml is my $SGMLHOME)

```
CATALOG "/usr/local/sgml/openjade-1.3/dsssl/catalog"
```

```
CATALOG "/usr/local/sgml/dsssl/docbook/catalog"
CATALOG "/usr/local/sgml/docbook/4.1sgml/catalog"
CATALOG "/usr/local/sgml/docbook/4.1.2xml/catalog"
SYSTEM "http://www.oasis-open.org/docbook/xml/4.1.2/docbookx.dtd"
"docbook/4.1.2xml/docbookx.dtd"
```

The CATALOG entries list catalogs to look through. The SYSTEM entry maps the SYSTEM id, http://www.oasis-open.org/docbook/xml/4.1.2/docbookx.dtd to the DTD on our local system (also a SYSTEM id). I *believe* that this is because openjade/jade can't retrieve from URLs. Anybody have any input?

One quirk(?) of openjade is that it doesn't support DTDDECL in catalog files. Your error logs will contain LOTS of lines like this.

```
openjade:/usr/local/sgml/docbook/4.1sgml/catalog:22:0:W: DTDDECL catalog entries are
not supported
openjade:/usr/local/sgml/docbook/4.1sgml/catalog:22:0:W: DTDDECL catalog entries are
not supported
openjade:/usr/local/sgml/docbook/4.1sgml/catalog:22:0:W: DTDDECL catalog entries are
not supported
```

It is normal and acceptable, but there is a work around I've seen referenced. To remove this annoying message, Edit the Docbook SGML 4.1 catalog file. $SGMLHOME/docbook/4.1sgml/catalog. Comment out this line:

```
DTDDECL "-//OASIS//DTD DocBook V4.1//EN" "docbook.dcl"
```

By adding two -- characters before and after as shown below. This comments out the line.

```
-- DTDDECL "-//OASIS//DTD DocBook V4.1//EN" "docbook.dcl"  --
```

Make sure that you add $SGMLHOME/openjade-1.3/bin to your path and that you update your environment so that the loader can find the libraries located in $SGMLHOME/openjade-1.3/lib. I edited /etc/profile and /etc/ld.so.conf; YMMV.

5. Using the Tools

This is a terse introduction to using the DocBook tools to compile XML documents. I won't go into the details of DocBook mark up. See DocBook: The Definitive Guide (http://docbook.org/tdg/en/html/docbook.html) for complete information on writing DocBook markup.

5.1. A Simple DocBook XML document

 Below is an example of a very simple DocBook XML document. Copy it into a text file and we'll compile it into HTML and RTF in a minute.

text.xml (sample/test.xml) [as text (sample/test.xml.txt)]

```
<?xml version="1.0" encoding='ISO-8859-1'?>
<!DOCTYPE book PUBLIC "-//OASIS//DTD DocBook XML V4.1.2//EN"
                      "http://www.oasis-open.org/docbook/xml/4.1.2/docbookx.dtd"

[
<!ENTITY version "0.01">
<!ENTITY supercopy "SUPERCOPYRIGHT &copy;&reg;">
]>


<!--
Comments just like html
-->

<book>

  <bookinfo>
    <title>Simple XML Sample Document</title>
    <author>
      <firstname>John</firstname>
      <surname>Doe</surname>
    </author>
    <copyright><year>2001</year><holder>John Doe</holder></copyright>
    <legalnotice>
      <para>This legal mumbo jumbo will stop evil. It even has &supercopy;.</para>
    </legalnotice>
    <abstract>
      <para>This is a simple XML sample version &version;. It is good for nothing but
processing.</para>
    </abstract>
  </bookinfo>

  <chapter>
    <title>About this book</title>
    <para>
      This book was hard
      work if you look at the history and momentum behind the LDP and
      DocBook and SGML and XML and....and....and...</para>

    <section><title>Copyrights and Trademarks</title> <para>Copyright
        &copy; 2001 John Doe</para>
    </section>

    <section><title>Purpose/Scope</title>
      <para>This guide is tightly scoped with one purpose; to process.</para>
    </section>

  </chapter>

  <appendix><title>References</title>
    <para>Some Hoity Toity Person</para>
```

```
    </appendix>

</book>
```

5.2. Generating HTML & RTF

You can see that DocBook markup is pretty straight forward. Now let's compile this test.xml into some real LDP HOWTO style HTML. I made a directory called 'sample' to work with.

```
mkdir sample
cd sample
mv $DBARCHIVE/test.xml .
openjade -t xml -d $SGMLHOME/dsssl/docbook/html/ldp.dsl#html
$SGMLHOME/dsssl/docbook/dtds/decls/xml.dcl test.xml
```

You shouldn't get any error messages. After, processing you should have the following (or similar) in the folder with test.xml

```
a23.html   index.html   test.xml   x20.html
c14.html   ln10.html    x17.html
```

To compile RTF try this

```
openjade -t rtf -d $SGMLHOME/dsssl/docbook/print/ldp.dsl#print
$SGMLHOME/dsssl/docbook/dtds/decls/xml.dcl test.xml
```

Description of command line switces:

- -t backend to use (fot|rtf|tex|mif|sgml|xml)
- -d style sheet to use

That's all there is to it. Once you figure out TeX and RTF you can convert to other popular document formats with other tools. You might use TeX, LaTeX, PDF, PDB, or even some unmentioned proprietary formats and tools. I've made my sample outputs available with this guide.

- Sample HTML output (sample/index.html)
- Sample RTF output (sample/test.rtf)

6. References

I glommed all this together from mailing lists and documents listed below.

- Linux Documentation Project (http://www.linuxdoc.org/)

- Authors Page (http://www.linuxdoc.org/authors/index.html#resources)
- LDP Author Guide (http://www.linuxdoc.org/LDP/LDP-Author-Guide/index.html)
- The docbook mailing list (http://www.linuxdoc.org/mailinfo.html) (Thanks to Greg Fergusson and Dan Scott for their suggestions)
- LDP Style Sheet (http://www.linuxdoc.org/authors/tools/ldp.dsl)

- Godoy's Docbook Page (http://www.ibiblio.org/godoy/)
- DocBook Install mini-HOWTO (http://www.linuxdoc.org/HOWTO/mini/DocBook-Install/)
- DocBook: The Definitive Guide by Norman Walsh (http://docbook.org/tdg/en/html/docbook.html)
- Setup Instructions For A Coherent SGML/DocBook Environment (http://adl.opengroup.org/exgr/papers/sgml_setup.html)
- Standard Deviations from Norm: If You Can Name It, You Can Claim It! (Norm Walsh article) (http://www.oasis-open.org/cover/walshSystemsIDsNewURL.html)
- A gentle guide to DocBook (IBM article) (http://www-106.ibm.com/developerworks/library/l-docbk.html?dwzone=linux)
- Making PDF documents with DocBook (An incredibly helpful article) (http://linuxfocus.saneg.itu.edu.tr/English/May2000/article152.shtml)
- OpenJade home (http://openjade.sourceforge.net)
- DocBook Technical Commitee (http://www.oasis-open.org/docbook/)
- DocBook Home at Sourceforge (http://docbook.sourceforge.net)
- Jim Weller's sleepless notes (http://www.jimweller.net) (not pretty)
  - Title Bout: Jim v. Doctools (http://www.jimweller.net/article.php?sid=26)
  - Using Doctools XML (http://www.jimweller.net/article.php?sid=27)