Really quick guide to DocBook

Ferry Boender

1. Introduction

This document is about DocBook. DocBook is a standard for creating, mostly technical, documents. DocBook's great advantage lies in the fact that it allows you to convert one source format into multiple target formats. I.e. one SGML document can be converted into HTML, Postscript, PDF, RTF, etc. This document will briefly touch upon a number of subjects having to do with DocBook. How to install DocBook and conversion tools, how to write DocBook SGML and how to convert them into a useful readable document.

First off I'd like to say that I am by no means a DocBook expert. I am, probably just like you, a complete newbie. So why did I write this small guide than? Because I'm presumably one, perhaps two, steps ahead of the people who have never dealt with DocBook at all. It can be a bit of a pain in the back in order to get started with DocBook, and I know nobody likes to read through a 700 page book about the subject. So there, that's why I wrote this. But because I'm a newbie too, don't trust anything I say. If stuff doesn't work, check the books/manuals/dtd's/someone with a clue.

Second of all, this document describes only DocBook SGML, not DocBook XML. If you have absolutely no clue about what this means, you should really take the time to read the section about 'What is DocBook?' and 'What is SGML?'. This document also is, and will remain for some time, a work in progress. My hope is that other people will also contribute to this document. This might be in vain though.

2. About DocBook

2.1. What is DocBook?

DocBook is a standard. It's a set of rules that explain what you can and can't do in a document. For instance, it will tell you (or rather some tool which will check up on you to see if you're sticking to the rules) that you can't put an computer program example in an address specification. Or you can't put text outside a paragraph. Those kinds of things.

Now this DocBook standard is defined as an SGML DTD. That's right, an SGML DTD.

2.2. What is SGML?

If you know what XML is all about then, well, SGML is almost exactly the same. There actually also is a DocBook XML version. If you don't know diddlysquat about XML, read on. SGML is a means to define rules. SGML, an abbreviation of Standard Generalized Markup Language, allows these rules to define what can and can't be done in a markup language document (like, i.e. HTML). These rules are placed in a document which is called a DTD (Document Type Definition). Through the use of this DTD, tools can check if everything you

wrote in the marked up document is valid according to the rules defined in that DTD. That is all that SGML and DocBook are.

So neither SGML nor the DocBook DTD can actually convert your source SGML document to anything else! All they do is define rules. Some tool somewhere does all the real work, converting your documents to what you want. So essential this isn't just a DocBook guide, but also a guide to DocBook converting tools.

2.3. Getting and installing DocBook

Getting and installing DocBook is actually very simple. You just download the DTD document, and you're done. But of course you also want some tools and such accompanying the DTD so you can actually do something useful with it. Information about various tools can be found in the 'Tools' section.

3. DocBook SGML documents

SGML documents are a lot like HTML. (I'm gonna get shot, in a SGML/XML-expert-driveby-shooting now. You're not allowed to say this I believe). They contain tags (called elements in SGML and XML), which in turn contain attributes. Composing a SGML document is the same as a HTML document. You put some tags here, some tags there, some tags everywhere. The big difference with DocBook SGML and HTML is that HTML defines what the target document will actually look like (font size = 3), while the DocBook SGML just defines how the document is built up. So instead of saying: This piece of text, which is the title of the document, should look large, bold and slanted, the DocBook SGML just says: This is the document's title. They leave it up to the conversion tools to decide how the document will eventually look.

3.1. Some basics

Writing a DocBook document starts with marking the document as such:

```
<!DOCTYPE article PUBLIC "-//OASIS//DTD DocBook V4.1//EN">
```

Translation: This document is of the type 'article', and uses OASIS' DTD definition version v4.1. The 'article' part may vary. You can change it to 'book' in order to create a book and all kinds of other things. The exact difference between these, and its effect on the conversion tools is unknown to me.

Above line should be the very first thing you put in your (plain text) document. After that you specify the global layout of your document. This is done with a tag which has the same name as the DOCTYPE you used on the first line. In this case 'article'. This is what the basic DocBook document looks like:

```
<!DOCTYPE article PUBLIC "-//OASIS//DTD DocBook V4.1//EN">
<article lang="en">
```

```
    <!-- some stuff will go here -->
</article>
```

If you were making a book, the first line would read 'book' instead of 'article' and the start and end element which are now known as '<article>' would be '<book>'. The entire contents of your article/book/whatevah will be placed between these two elements. Now, you can't just dump some text in between them. Why not? Cause the rules of DocBook say you can't, that's why. What can you put between the elements than? Well, lots of things actually. (check the 'Additional reading' section). For now I'll present you with this:

```
<sect1>
<para>Never let school interfere with your education.</para>
</sect1>
```

Two new things there: 'sect1' and 'para'. Let's guess they stand for 'Section 1' and 'Paragraph'. A section is exactly which it says it is: It's a section of your article. There are also those things like : <sect1> <sect2> (which can be nested inside each other to give your document a hierarchy) <chapter>, etc.

3.2. A simple DocBook document

You should now have enough information to create a simple SGML document. Here it goes:

```
<!DOCTYPE article PUBLIC "-//OASIS//DTD DocBook V4.1//EN">
<article lang="en">

<sect1>
    <title>Introduction</title>
    <para>Welcome to my very first introduction to DocBook!</para>
</sect1>
</article>
```

That's it! Now, read the section about 'Converting DocBook' so you can see some output documents. For those of you who are impatient UNIX users, issue the following command:

[todsah@eek]~$ **db2html -u my_first_db.sgml**

3.3. Useful DocBook elements

DocBook elements are the key to it being of any use. So it sure be pleasant if you'd know some of them. Here are some of the most used ones. Most used by me that is. For more, check out the 'Additional readings' section.

3.3.1. Layout elements

`<sect1>` (http://www.docbook.org/tdg/en/html/sect1.html)

> sect elements define the overall layout of you document. They will be automatically numbered (depending on the tool used to convert it, of course) in sequence. You can nest them, that is, a `<sect2></sect2>` can be nested within a `<sect1>` element.

`<para>` (http://www.docbook.org/tdg/en/html/para.html)

> A paragraph. It can contain normal text.

`<title>` (http://www.docbook.org/tdg/en/html/title.html)

> Some elements are allowed to be titled. Most notably, the `<sect#>` elements and the `<articleinfo>` (http://www.docbook.org/tdg/en/html/articleinfo.html) element.

`<articleinfo>` (http://www.docbook.org/tdg/en/html/articleinfo.html)

> Guess what? It allows for article information to be described. It can contain a couple of elements, namely: `<title>` (http://www.docbook.org/tdg/en/html/title.html), `<author>` (http://www.docbook.org/tdg/en/html/author.html) and some more.

`<author>` (http://www.docbook.org/tdg/en/html/author.html)

> An element to hold other elements which describe something about the author. `<firstname>` (http://www.docbook.org/tdg/en/html/firstname.html) and `<surname>` (http://www.docbook.org/tdg/en/html/surname.html) would look nice within this element.

`<screen>` (http://www.docbook.org/tdg/en/html/screen.html)

> The screen element should contain elements which describe a computer screen. These are, amongst others, `<computeroutput>` (http://www.docbook.org/tdg/en/html/computeroutput.html) and `<userinput>` (http://www.docbook.org/tdg/en/html/userinput.html)

`<computeroutput>` (http://www.docbook.org/tdg/en/html/computeroutput.html)

> This element's contents (text) represents information which is shown on a computer screen as output by the computer.

`<userinput>` (http://www.docbook.org/tdg/en/html/userinput.html)

> This element's contents (text) represents information which is typed on the keyboard by the user.

`<itemizedlist>` (http://www.docbook.org/tdg/en/html/itemizedlist.html)

> Itemized lists are those nice lists which have bulletpoints in front of them. How they are represented once converted to an output document depends on the tool used, but usually it will be bulletpoints. The itemizedlist element should contain `<listitem>` (http://www.docbook.org/tdg/en/html/listitem.html) elements.

`<listitem>` (http://www.docbook.org/tdg/en/html/listitem.html)

> This element's contents, which should be `<para>` (http://www.docbook.org/tdg/en/html/para.html) (which in turn will contain some text), represents one item in an `<itemizedlist>`.

`<variablelist>` (http://www.docbook.org/tdg/en/html/variablelist.html)

> A variable list is like an itemized list, except it allows you to define a term and a description for this term. It's structure is somewhat more difficult than any other elements described in this document. The variablelist element should contain `<varlistentry>` elements.

`<varlistentry>` (http://www.docbook.org/tdg/en/html/varlistentry.html)

> A `<variablelist>` (http://www.docbook.org/tdg/en/html/variablelist.html) entry. This element should contain a `<term>` (http://www.docbook.org/tdg/en/html/term.html) element and a `<listitem>` (http://www.docbook.org/tdg/en/html/listitem.html) element per `<varlistentry>` (http://www.docbook.org/tdg/en/html/varlistentry.html) element.

`<term>` (http://www.docbook.org/tdg/en/html/term.html)

> Contains, in text, the term which will be described.

`<programlisting>` (http://www.docbook.org/tdg/en/html/programlisting.html)

> This should contain source code for programs and examples.

4. Converting DocBook

As I said in the introduction, the real power of DocBook is the ability to convert DocBook documents to any number of output formats using tools. Using these tools it's possible to convert a SGML source document to PDF, HTML, RTF, and more.

There are also other tools available which will convert documents to DocBook instead of the other way around. Tools for checking the syntax, tidying up the source, editors, etc. This chapter will only discuss some of the conversion tools.

4.1. Tools

Various conversion tools exist, some more used than others. Some of the most used are:

Unix:

> OpenJade (http://openjade.sourceforge.net/)

> OpenJade is a tool for converting (publishing) DocBook documents to HTML, PDF, RTF and other formats. It is an implementation of the DSSSL standard, which is a fancy acronym for a set of rules which can be applied to DocBook documents in order to convert them. OpenJade is Free Software, and can be freely downloaded from the above page. Redhat users can install the `openjade-*` RPM packages. Debian users can install the `docbook`, `docbook-dsssl`, `docbook-utils` and `openjade1.3` packages (apt-get powah)

Windows:

> I haven't tried out DocBook on the Windows platform. You're on your own. If you know about how to do this, please email me so I can include the information here.

4.2. Converting

On UNIX, converting is easy. Just switch to the directory containing your DocBook source, and issue the following command:

db2html my_first_db.sgml

You can also use db2pdf, db2ps or db2rtf to convert your DocBook to PDF, Postscript or Rich Text Format. The db2 commands are easy to use wrappers to OpenJade and are part of the JW tool-set.. They are available in the docbook-utils debian package.

I don't know anything about publishing DocBook on the Windows (or any other) platform. If you know something about this subject, please email it to me.

4.3. Configuring tools

The OpenJade tool can be configured by altering the `dbparams.dsl` files, which can be found in the `print/` and `html/` directories in the `stylesheet/dsssl/modular/` directory. (This directory may vary from system to system. My Debian machine has them located in the `/usr/share/sgml/docbook/stylesheet/dsssl/modular/` directory) The `html` directory contains stylesheets for the conversion of DocBook SGML to HTML. The `print` directory contains stylesheets for the conversion to various printable formats like PDF and RTF.

For now, the only useful option I will discuss is one in the `html/dbparams.dsl` file.

4.3.1. HTML: Setting a Cascading StyleSheet (CSS)

CSS are stylesheets which determine how the different HTML elements will eventually look when displayed in the web browser. Please note that they have almost absolutely nothing to do with DSSSL stylesheets. If you want the generated HTML to contain some kind of CSS file, you can specify one. This file will then be referenced by the produced HTML. The option you should be looking for is the %stylesheet% option. Initially, it looks like this:

```
(define %stylesheet%
;; REFENTRY stylesheet
;; PURP Name of the stylesheet to use
;; DESC
;; The name of the stylesheet to place in the HTML LINK TAG, or '#f' to
;; suppress the stylesheet LINK.
;; /DESC
;; AUTHOR N/A
```

```
;; /REFENTRY
#f)
```

The '#f' indicates that no CSS stylesheet should be used. if you want to change it, you can specify a stylesheet like this:

```
(define %stylesheet% "http://todsah.nihilist.nl/css/docbook_fb.css")
```

This will generate the following additional HTML code in the output of openjade's DocBook to HTML conversion:

```
<META
NAME="GENERATOR"
CONTENT="Modular DocBook HTML Stylesheet Version 1.7"><LINK
REL="STYLESHEET"
TYPE="text/css"
HREF="http://todsah.nihilist.nl/css/docbook_fb.css">
```

As you can see, the stylesheet is included in the HTML output. You can also see that the HTML output of openjade is quite ugly. Just ignore it.

5. Additional readings

This document just barely scratches the surface of DocBook, SGML, jade/openjade, DSSSL and all other things DocBook. For more information, you might like to consult these additional readings. I've tried to stick with online resources so you won't have to leave the comfort of your home in order to get DocBook literate. Two of the most important resources you should know about are these:

DocBook: The Definitive Guide
(http://www.oasis-open.org/docbook/documentation/reference/html/docbook.html)

> ISBN: 156592-580-7

> DocBook: The definitive guide is pretty close to being definitive. It contains just about everything you need or want to know about DocBook SGML/XML. Also includes instructions on various Windows/Mac/Unix tools for DocBook. The book can be bought, but for the freeloaders among us there is an free online version available. You can reach it by following the link above.

Docbook Frequently Asked Questions (http://www.dpawson.co.uk/docbook/)

> This web-page covers frequently asked questions about DocBook, and it is a good place to start if you're new to DocBook. It also shows that DocBook: The definitive guide isn't as definitive as they'd make you believe. This is because the writers of the book forgot an element reference guide. Luckily, DocBook FAQ page has one available here (http://www.dpawson.co.uk/docbook/reference.html#d12e58).

6. About this document

6.1. Contributing

This document is incomplete. For instance, it only talks about UNIX tools for converting DocBook. If you'd like to contribute something to this document, like i.e. extra elements, other tools, extra information, spelling corrections, etc, you may send them to me (consult the author section) and I will include them in this document.

6.2. Copyright

Copyright (c) 2002-2004, Ferry Boender

This document may be freely distributed, in part or as a whole, on any medium, without the prior authorization of the author, provided that this Copyright notice remains intact, and there will be no obstruction as to the further distribution of this document. You may not ask a fee for the contents of this document, though a fee to compensate for the distribution of this document is permitted.

Modifications to this document are permitted, provided that the modified document is distributed under the same license as the original document and no copyright notices are removed from this document. All contents written by an author stays copyrighted by that author.

Failure to comply to one or all of the terms of this license automatically revokes your rights granted by this license

All brand and product names mentioned in this document are trademarks or registered trademarks of their respective holders.

6.3. Author

Ferry Boender
```
<f (DOT) boender (AT) electricmonk (DOT) nl>
```

6.4. History

This document has undergone the following changes:

Revision History

Revision 0.1 July 27, 2003 Revised by: FB
Initial release.