
DocBook V5.0

The Transition Guide

24 October 2005

This version:

<http://docbook.org/docs/howto/2005-10-24/>

Latest version:

<http://docbook.org/docs/howto/>

Previous version:

<http://docbook.org/docs/howto/2005-10-21/>

Editors:

Jirka Kosek, <jirka@kosek.cz>

Norman Walsh, <ndw@nwalsh.com>

Table of Contents

Introduction	1
Finally in a namespace	2
Relaxing with DocBook	2
Why switch to DocBook V5.0?	3
Schema jungle	3
Toolchain	4
Editing DocBook V5.0	4
Validating DocBook V5.0	9
Processing DocBook V5.0	10
Markup changes	11
Improved cross-referencing and linking	11
Renamed elements	12
Removed elements	13
Converting DocBook V4.x documents to DocBook V5.0	13
What About Entities?	13
Customizing DocBook V5.0	14
FAQ	14
Bibliography	15

This document is targeted at DocBook users who are considering switching from DocBook V4.x to DocBook V5.0. It describes differences between DocBook V4.x and V5.0 and provides some suggestions about how to edit and process DocBook V5.0 documents. There is also section devoted to conversion of legacy documents from DocBook 4.x to DocBook V5.0.

Introduction

The differences between DocBook V4.x and V5.0 are quite radical in some aspects, but the basic idea behind DocBook is still the same and almost all element names are unchanged. Because of this it is very easy to become familiar with DocBook V5.0 if you know any previous version of DocBook. You can find complete list of changes in [DB5SPEC], here we will discuss only the most fundamental changes.

Finally in a namespace

All DocBook V5.0 elements are in the namespace `http://docbook.org/ns/docbook`. XML namespaces are used to distinguish between different element sets. In the last few years, almost all new XML grammars have used their own namespace. It is easy to create compound documents that contain elements from different XML vocabularies. DocBook V5.0 is following this design rule. Using namespaces in your documents is very easy. Consider this simple article marked up in DocBook V4.5:

```
<article>
  <title>Sample article</title>
  <para>This is a really short article.</para>
</article>
```

The corresponding DocBook V5.0 article will look very similar:

```
<article xmlns="http://docbook.org/ns/docbook" ...>
  <title>Sample article</title>
  <para>This is a really short article.</para>
</article>
```

The only change is the addition of a default namespace declaration (`xmlns="http://docbook.org/ns/docbook"`) on the root element. This declaration applies the namespace to the root element and all nested elements. Each element is now uniquely identified by its local name and namespace.

Note

The namespace name `http://docbook.org/ns/docbook` serves only as an identifier. This resource is not fetched during processing of DocBook documents and you are not required to have an Internet connection during processing. If you access the namespace URI with a browser, you will find a short explanatory document about the namespace. In the future this document will probably conform to (some version of) RDDL and provide pointers to related resources.

Relaxing with DocBook

For more than decade, the DocBook schema was defined using a DTD. However DTDs have serious limitations and DocBook V5.0 is thus defined using a very powerful schema language called RELAX NG. Thanks to RELAX NG, it is now much easier to create customized versions of DocBook and some content models are now cleaner and more precise.

Using RELAX NG has an impact on the document prolog. The following example shows the typical prolog of a DocBook V4.x document. The version of the DocBook DTD (in this case 4.5) is indicated in the document type declaration (!DOCTYPE) which points to a particular version of the DTD.

Example 1. DocBook V4.5 document

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE article PUBLIC "-//OASIS//DTD DocBook XML V4.5//EN"
    'http://www.oasis-open.org/docbook/xml/4.5/docbookx.dtd'>
<article lang="en">
  <title>Sample article</title>
  <para>This is really very short article.</para>
</article>
```

In contrast, DocBook V5.0 does not depend on DTDs anymore. This means that there is no document type declaration and the version of DocBook used is indicated with the `version` attribute instead.

Example 2. DocBook V5.0 document

```
<?xml version="1.0" encoding="utf-8"?>
<article xmlns="http://docbook.org/ns/docbook" version="5.0" xml:lang="en">
  <title>Sample article</title>
  <para>This is really very short article.</para>
</article>
```

As you can see, DocBook V5.0 is built on top of existing XML standards as much as possible and the `lang` attribute is superseded by the standard `xml:lang` attribute.

Another fundamental change here is that there is no direct indication of the schema used. Later in this document, you will learn how you can specify a schema to be used for document validation.

Note

Although using the RELAX NG schema with DocBook V5.0 is recommended, there are also DTD and W3C XML Schema versions available (see the section called “Where to get the schemas”) to satisfy tools that do not yet support RELAX NG.

Why switch to DocBook V5.0?

The simple answer is “because DocBook V5.0 is the future”. Apart from this marketing blurb, there are also more technical reasons:

- *DocBook V4.x is feature frozen.* At the time of this writing DocBook V4.5 is the last version of DocBook in V4.x series. Any new DocBook development, like the addition of new elements, will be done in DocBook V5.0. It is only a matter of time before useful, new elements will be added into DocBook V5.0, but they are not likely to be back ported into DocBook V4.x. DocBook V4.x will be developed in a maintenance mode and errata will be published if necessary.
- *DocBook V5.0 offers new functionality.* Even the current version of DocBook V5.0 provides significant improvements over DocBook V4.x. For example there is general markup for annotations, a new and flexible system for linking, ...
- *DocBook V5.0 is more extensible.* Having DocBook V5.0 in a separate namespace allows you to easily mix DocBook markup with other XML based languages like SVG, MathML, XHTML or even FooBarML.
- *DocBook V5.0 is easier to customize.* RELAX NG offers many powerful constructs that make customizing the existing DocBook schema very easy. Now it is much easier to create a customized DocBook version than before with DTDs.

Schema jungle

Schemas for DocBook V5.0 are available in several formats at <http://www.oasis-open.org/docbook/xml/5.0b1/> (or the mirror at <http://docbook.org/xml/5.0b1/>). Only the RELAX NG schema is normative and it is preferred over the other schema languages. However, for your convenience there are also DTD and W3C XML Schema versions provided for DocBook V5.0. But please note that neither DTDs nor XML schemas are able to capture all the constraints of DocBook V5.0. This means that a document that validates against the DTD or XML schema is not necessarily valid against RELAX NG schema and thus can't be considered a valid DocBook V5.0 document.

DTD and W3C XML Schema versions of the DocBook V5.0 grammar are provided as a convenience for users who want to use DocBook V5.0 with legacy tools that don't support RELAX NG. Authors are encouraged to switch to RELAX NG based tools as soon as possible, or at least to validate documents against the RELAX NG schema before further processing.

Where to get the schemas

The latest versions of schemas can be obtained from the following locations:

RELAX NG schema	http://www.oasis-open.org/docbook/xml/5.0b1/rng/docbook.rng
RELAX NG schema in compact syntax	http://www.oasis-open.org/docbook/xml/5.0b1/rng/docbook.rnc
DTD	http://www.oasis-open.org/docbook/xml/5.0b1/dtd/docbook.dtd
W3C XML Schema	http://www.oasis-open.org/docbook/xml/5.0b1/xsd/docbook.xsd

These schemas are also available from the mirror at <http://docbook.org/xml/5.0b1/>.

DocBook documentation

Detailed documentation about each DocBook V5.0 element is presented in the reference part of *DocBook: The Definitive Guide* [<http://docbook.org/tdg5/en/html/pt02.html>].

Note

Other parts of the book have not yet been updated to reflect the changes made in DocBook V5.0. Please do not be confused by this.

Toolchain

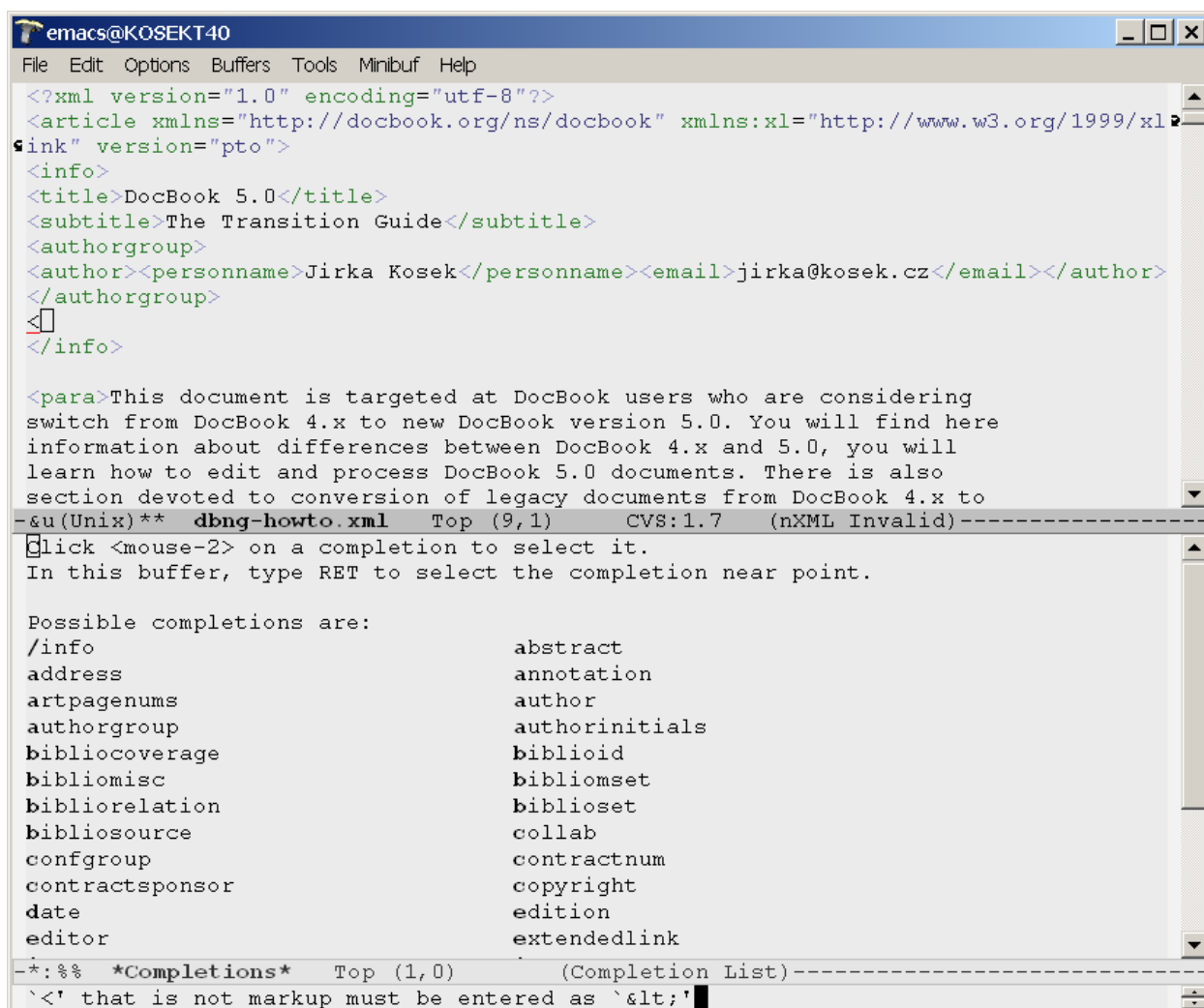
There are a lot of questions concerning tools that are able to work with DocBook V5.0 documents. The aim of this sections is to briefly describe the tools and procedures that should be used to edit and process content stored in DocBook V5.0.

Editing DocBook V5.0

Because DocBook is an XML based format and XML is a text based format, you can use any text editor to create and edit DocBook V5.0 documents. However using “dumb” editors like Notepad is not very productive. You will do much more better if you use editor that comes with some XML support. As there are DTD and W3C XML Schemas available for DocBook V5.0, you can configure your favorite editor to use these schemas. But as we said already, it is recommended that you use the RELAX NG grammar with DocBook V5.0. The rest of this section contains an overview of XML editors (listed in alphabetical order) that are known to support guided editing based on RELAX NG schema.

Emacs and nXML

nXML mode [<http://www.thaiopensource.com/nxml-mode/>] is an addon for the GNU Emacs text editor. By installing nXML you can turn Emacs into a very powerful XML editor which will offer guided editing and validation of XML documents.

Figure 1. Emacs with nXML mode is providing guided editing and validation

nXML uses special configuration file named `schemas.xml` to associate schemas with XML documents. Often you will find this file in the directory `site-lisp/nxml/schema` inside the Emacs installation directory. Adding the following line into the configuration file, will associate DocBook V5.0 elements with the appropriate schema:

```
<namespace ns="http://docbook.org/docbook-ng" uri="/path/to/docbook.rnc"/>
```

Note

Please note that nXML ships with a file named `docbook.rnc`. This file contains the RELAX NG grammar for DocBook V4.x. Be sure that you associate DocBook V5.0 namespace with the corresponding DocBook V5.0 grammar.

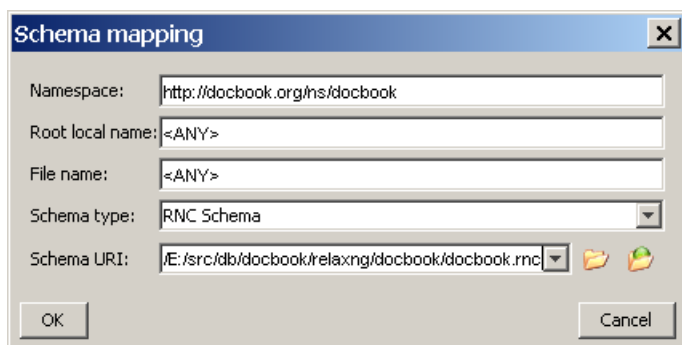
If you can't edit global `schemas.xml` file, you can create this file in a directory with your document. nXML will find associations placed here also. In this case you must create complete configuration file like:

```
<locatingRules xmlns="http://thaiopensource.com/ns/locating-rules/1.0">
  <namespace ns="http://docbook.org/docbook-ng" uri="/path/to/docbook.rnc"/>
</locatingRules>
```

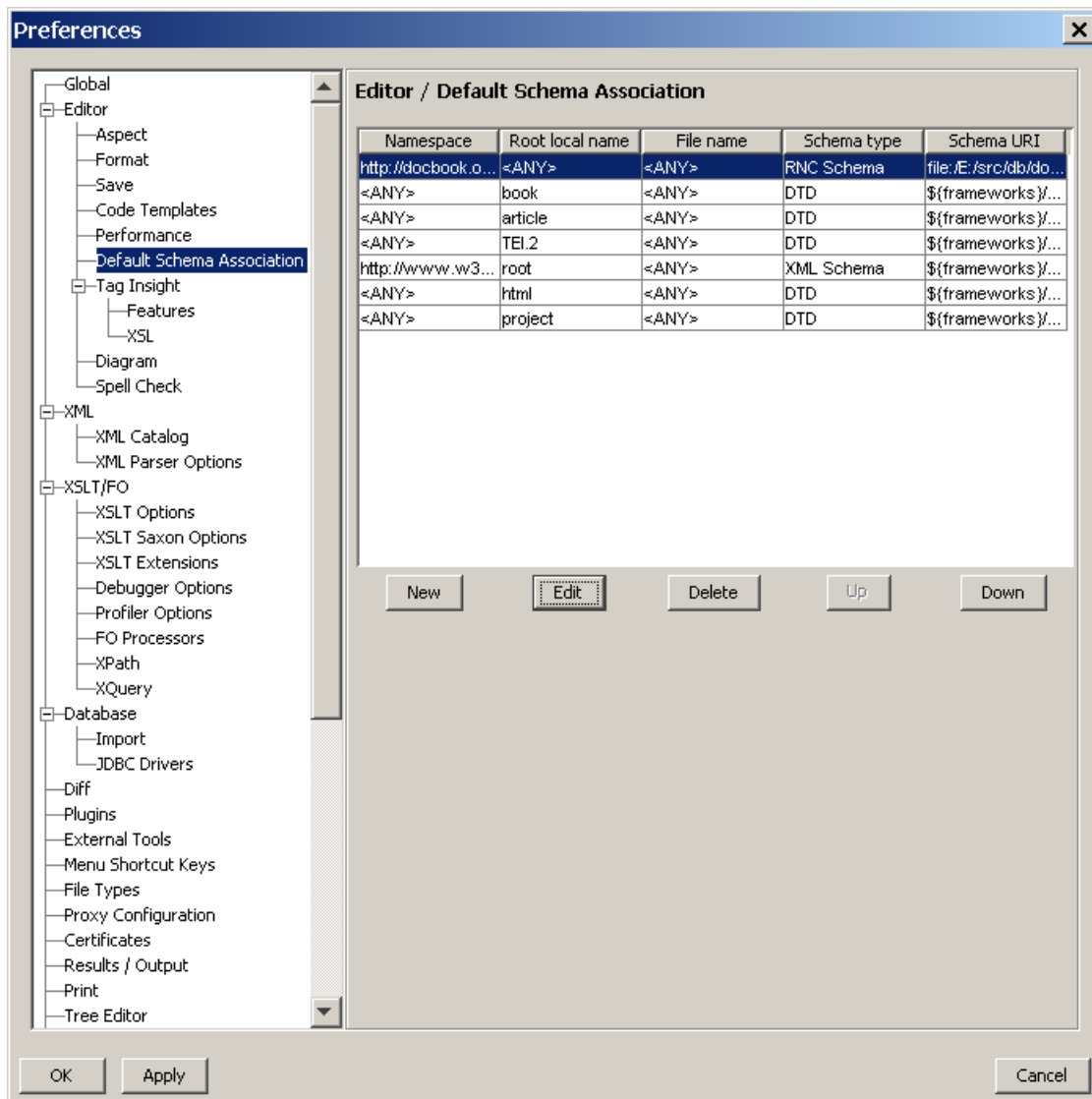
oXygen

oXygen is a feature rich XML editor. It has built in support for many schema languages including RELAX NG. If you want to smoothly edit and validate DocBook 5.0 documents you should associate DocBook namespace with the corresponding schema. Go to Options → Preferences... → Editor → Default Schema Associations. Then click New button to add new association. Type in the DocBook namespace, RELAX NG compact syntax schema location and choose appropriate type of schema and confirm it by pressing OK.

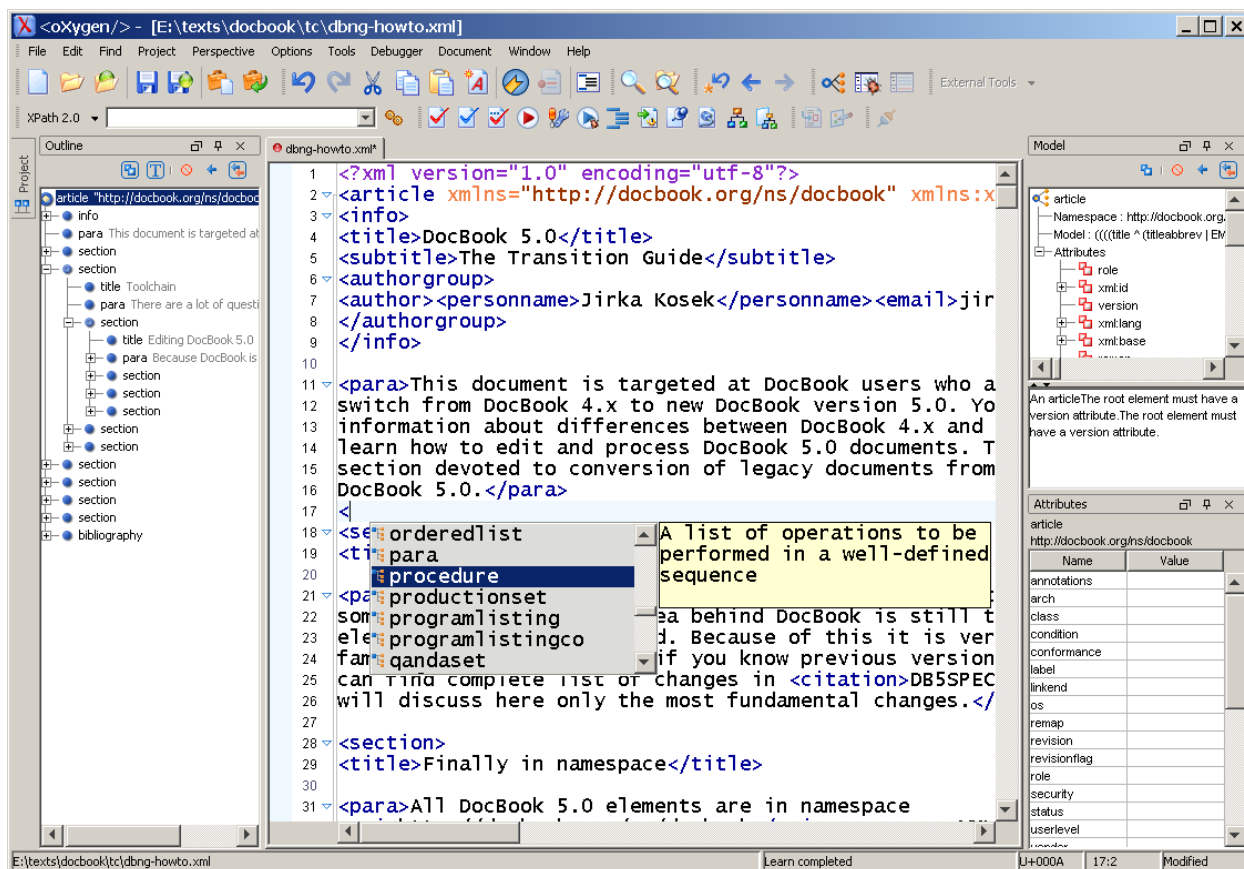
Figure 2. Adding new schema association in oXygen



Because oXygen comes with some preconfigured associations for DocBook V4.x, you must move your newly added one to the top of the list (using Up button). That way you will be able to use oXygen with both DocBook V4.x and DocBook V5.0.

Figure 3. DocBook V5.0 association must precede associations for DocBook V4.x

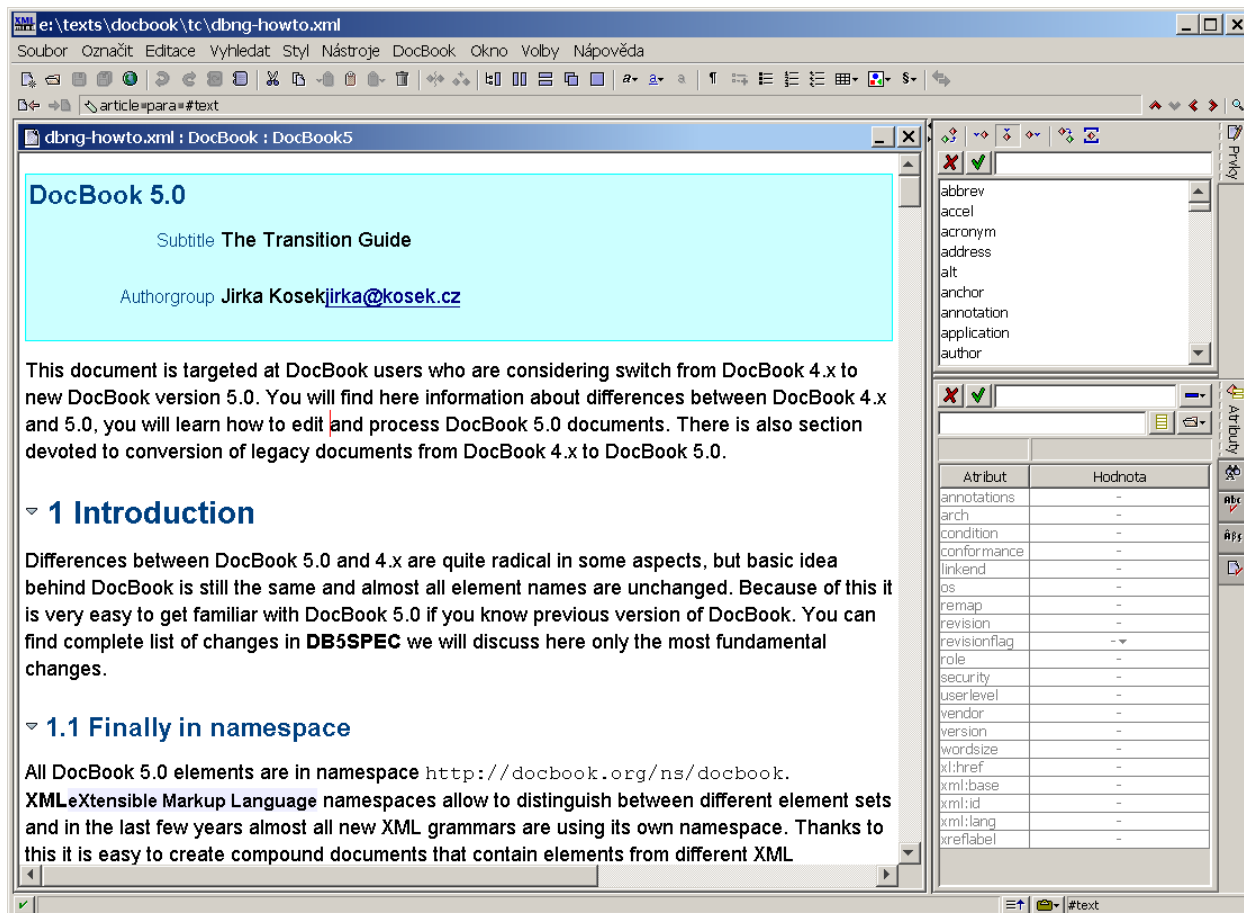
Now you can close preference by clicking on OK button.

Figure 4. DocBook V5.0 document opened in oXygen

XML Mind XML editor

XML Mind XML editor (XXE) is a visual validating XML editor that provides word-processor like interface to users. It is available in two versions—Standard and Professional. Standard version is free and provides everything you need to edit DocBook V5.0 documents.

Figure 5. XML Mind XML Editor – feels almost like MS Word but real DocBook V5.0 markup is created



Since version 2.11, XXE comes with bundled DocBook V5.0 configuration. Unfortunately this configuration is not enabled by default. You must copy content of the directory `XXE_install_dir/doc/rnsupport/config/docbook5/` into `XXE_install_dir/addon/config/docbook5/` in order to activate it. After restart of XXE you will be able to create (template for article is provided) and edit DocBook V5.0 documents.

RELAX NG schema provided with XXE can be outdated. If you want to use XXE with the latest schema just grab fresh copy of the `docbook.rng` and copy it over `XXE_install_dir/addon/config/docbook5/docbook.rng`.

Validating DocBook V5.0

If you are not using RELAX NG based validating editor when you are creating documents, it is highly recommended that you validate your documents before processing them. Only after successful validation you can be sure that your document is really DocBook V5.0 and that processing tools will be able to process it correctly.

You can find list of RELAX NG validators at <http://relaxng.org/#validators>. It is recommended to use validators with support for embedded Schematron rules inside RELAX NG schema. Schematron is a rule based validation language which is used to impose additional constraints on DocBook document. Schematron rules assert very subtle conditions which can not be expressed in a pure RELAX NG.

Sun Multi-Schema XML Validator (MSV) is able to validate XML document against RELAX NG and Schematron at the same time. To install and use MSV follow these steps:

1. Download file `relames.zip` from <https://msv.dev.java.net/servlets/ProjectDocumentList?folderID=101>.
2. Unpack the downloaded file into arbitrary directory.
3. Validate your document by using the following command:

```
java -Xss256K -jar /path/to/relames.jar /path/to/docbook.rng document.xml
```

Note

Switch `-Xss256K` is used to increase stack size of Java virtual machine. This is necessary because DocBook schema is quite large. If you get stack overflow errors from MSV, try to increase this value. If you are not using Java implementation from Sun, please consult documentation of your virtual machine to learn how to increase stack size.

Processing DocBook V5.0

Part of DocBook's great success can be attributed to the availability of free tools that can be used to transform DocBook content into various target formats including HTML and PDF. The DocBook XSL Stylesheets are among the most popular tools.

DocBook XSL Stylesheets

The DocBook stylesheets are written in a quite general way so that they have always been able to process content written in different versions of DocBook (for example 3.1 and 4.2). Recent versions of the stylesheets are also able to process DocBook V5.0 albeit with some limitations.

You can process DocBook V5.0 documents with the DocBook XSL stylesheets in exactly the same way as DocBook V4.x document. There is no need for a new special software, you can stick to you preferred XSLT processor, be it Saxon, xsltproc, Xalan or whatever else.

During document processing, the stylesheets are stripping namespaces from DocBook V5.0 in order to get document which will be very similar to DocBook V4.x. This is necessary because from the XSLT point of view elements from different namespaces are distinct and can not be easily processed by the same set of templates. This process is completely transparent to user. If you are processing DocBook V5.0 document with the stylesheets you will only see the following additional message:

```
Stripping NS from DocBook 5/NG document.
Processing stripped document.
```

Although you can successfully use existing stylesheets to process DocBook V5.0 there are some limitations. Some new features of DocBook V5.0 would require very complex rewrite of the stylesheets in order to support them. This is unlikely to happen because completely new version of stylesheets is currently being written from scratch. Examples of such unsupported features are:

- general annotations;
- general XLink links on all elements;

During namespace stripping, the base URI of the document is lost. This means that in rare situations, relatively referenced resources like images or program listings can be processed incorrectly. The stylesheets attempt to compensate for this problem, but it is possible that there are corner cases where it will fail.

XSLT 2.0 based reimplementation

XSLT 1.0 is missing some important features and, to overcome this limitation, the current DocBook XSL stylesheets use several implementation-specific extensions. Fortunately, authors of new XSLT version 2.0 were listening to these limitations and XSLT 2.0 adds many new and previously missing features into language. New XSLT 2.0 stylesheets are being implemented to process DocBook V5.0 documents with all new features.

Reimplementation of the stylesheets, and new functionalities of XSLT 2.0, allowed developers to integrate many new features into the stylesheets. Some of them are:

- seamless integration of profiling (conditional documents) with external bibliographies and glossaries;
- no need for (most) external extensions;
- internationalized indexes;
- easy to customize titlepage templates;

The disadvantage of XSLT 2.0 based stylesheets is that it is not finished yet and that there are not very many XSLT 2.0 implementations on the market. Currently the stylesheets are supporting only HTML and chunked HTML output. Other output formats are planned but do not expect them very soon because of limited free time of stylesheets developers.

But if you want to try the new stylesheets, of course you can. Just grab snapshot of development version of the stylesheets from <http://docbook.sourceforge.net/snapshots/docbook-xsl2-snapshot.zip> and unpack it somewhere. Then download and install Saxon 8 from <http://saxon.sf.net>.

To transform DocBook V5.0 document to a single HTML page you can then use command:

```
java -jar /path/to/saxon8.jar -o output.html document.xml /path/to/docbook-xsl2-snapshot/h
```

To transform DocBook V5.0 document to a set of chunked HTML pages you can then use command:

```
java -jar /path/to/saxon8.jar document.xml /path/to/docbook-xsl2-snapshot/html/chunk.xsl
```

Markup changes

You can find a complete list of changes in [DB5SPEC]. This section shows the most common markup changes between DocBook V4.x and V5.0 in several examples.

Improved cross-referencing and linking

In DocBook V4.x attribute `id` was used to assign unique identifier to element. In DocBook V5.0 this attribute is renamed to `xml:id` in order to comply to [XMLID].

Now you can use almost any inline element as a source of link, not just `xref` or `link`. So the following DocBook 4.x content:

```
<section id="dir">
  <title>DIR command</title>
  <para>...</para>
</section>

<section id="ls">
  <title>LS command</title>
```

```
<para>This command is synonymum for <link linkend="dir"><command>DIR</command></link> co
</section>
```

will be written in DocBook V5.0 as:

```
<section xml:id="dir">
  <title>DIR command</title>
  <para>...</para>
</section>

<section xml:id="ls">
  <title>LS command</title>
  <para>This command is synonymum for <command linkend="dir">DIR</command> command.</para>
</section>
```

Attribute `linkend` was added to all inline elements together with `href` attribute from XLink namespace. This means that you can use any inline element as a source of hypertext link. In order to use XLinks you have to declare XLink namespace (most often on the root element of your document):

```
<article xmlns="http://docbook.org/ns/docbook"
         xmlns:xl="http://www.w3.org/1999/xlink" version="5.0">
  <title>Test article</title>

  <para><application xl:href="http://www.gnu.org/software/emacs/emacs.html">Emacs</application>
    is my favourite text editor.</para>

  ...
```

Element `ulink` was completely removed from DocBook V5.0 in favor of XLink linking. Instead of DocBook V4.x `ulink` element:

```
<ulink url="http://docbook.org">DocBook site</ulink>
```

you can now use `link`

```
<link xl:href="http://docbook.org">DocBook site</link>
```

XLink links can contain a fragment identifier and you can even use them instead of `linkend` attributes to form cross-references inside document:

```
<command xl:href="#dir">DIR</command>
```

However XLink links are not checked during validation, while `xml:id/linkend` links are checked for ID/IDREF consistency. It depends on your needs what approach to internal linking is more suitable for you.

One case where the XLink-based, fragment identifier scheme is useful is when XInclude is being used. XML ID/IDREF links cannot span XInclude boundaries.

Renamed elements

Some elements were renamed to better express their meaning or to reduce total number of elements available in DocBook.

Table 1. Renamed elements

Old name	New name
sgmltag	tag
bookinfo, articleinfo, chapterinfo, *info	info
authorblurb	personblurb
collabname, corppauthor, corpcredit, corp-name	orgname
isbn, issn, pubsnumber	biblioid
lot, lotentry, tockback, tocchap, tocfrent, toclevel1, toclevel2, toclevel3, toclevel4, toclevel5, tocpart	tocdiv
graphic, graphicco, inlinegraphic, mediaobjectco	mediaobject and inlinemediaobject
ulink	link

Removed elements

The following elements were removed from DocBook V5.0 without any suitable replacement: `action`, `beginpage`, `highlights`, `interface`, `invpartnumber`, `medialabel`, `modespec`, `structfield`, `structname`.

Converting DocBook V4.x documents to DocBook V5.0

The DocBook V5.0 schema ships with an XSLT 1.0 stylesheet that is designed to transform valid DocBook V4.x documents to valid DocBook V5.0 documents.

To convert your document, `doc.xml` in the examples below, follow these steps:

1. Check the validity of your DocBook XML V4.x document. The conversion tool assumes that the input document is valid. If the input document contains markup errors, the results will be unpredictable at best.
2. Transform `doc.xml` to `newdoc.xml` with the `db4-upgrade.xsl` stylesheet included in the DocBook V5.0 distribution that you are using.
3. Check the validity of your DocBook XML V5.0 document against the DocBook V5.0 RELAX NG grammar.

In the vast majority of cases, the resulting document should be valid and your conversion process is finished.

If the document is not valid, please report the problem. (Over time, we'll have more experience with the sorts of things that can go wrong and we'll update this document to reflect that experience.)

What About Entities?

Using XSLT to transform existing documents to DocBook V5.0 has one potential disadvantage: it removes all entity references from your document.

If preserving entities is an important aspect of your production work flow, you will have to engage in a semi-manual process to preserve them.

1. Open your existing document using your favorite editing tool. You must use a tool that *is not* XML-aware, or one that allows you to edit markup “in the raw”.
2. Replace all occurrences of the entity references that you want to preserve with some unique string. For example, if you want to preserve “∏” references, you could replace them all with “[[[Product]]]” (assuming that the string “[[[Product]]]” doesn’t occur anywhere else in your document).
3. Copy the document type declaration off of your document and save it some place. The document type declaration is everything from “<!DOCTYPE” to the closing “[>”.
4. Perform the conversion described in the section called “Converting DocBook V4.x documents to DocBook V5.0”.
5. Open the new document using your favorite editing tool. Replace all occurrences of the unique string you used to save the entity references with the corresponding entity references.
6. Paste the document type declaration that you saved onto the top of your new document.
7. Remove the external identifier (the PUBLIC and/or SYSTEM keywords) from the document type declaration. A document that begins:

```
<!DOCTYPE book [  
<!ENTITY someEntity "some replacement text">  

```

is perfectly well-formed. If you don’t remove the references to the DTD, then your parser will likely try to validate against DocBook V4.0 and that’s not going to work. Alternatively, you could refer to the DocBook V5.0 DTD.

External Parsed Entities

External parsed entities, entities which load part of a document from another file, are a special case. These can often be replaced with XInclude elements.

The Perl script `db4-entities.pl`, also included in the DocBook V5.0 distribution attempts to perform this replacement for you. To use the script, perform the following steps:

1. Process your document with `db4-entities.pl`. The script expects a single filename and prints the XInclude version on standard output.
2. Process the XInclude version as described in the section called “Converting DocBook V4.x documents to DocBook V5.0”.

Customizing DocBook V5.0

TBD

FAQ

1. Authoring

- 1.1. How do I attach a schema to a DocBook V5.0 document when I do not want to use DTDs and !DOCTYPE?

There is no standard way of associating a RELAX NG schema with a document. Most tools provide some mechanism for performing this association, consult the documentation for your application. In some tools you must specify schema manually each time your want to edit/process your document.

1.2. How do I use entities like `–` in DocBook V5.0?

Modern schema languages (including RELAX NG and W3X XML Schema) do not provide any means to define entities that can be used for easier typing of special characters. Some editors provide functions or special toolbars that allow you to easily pick necessary character and insert it into document as a raw Unicode character or a numeric character reference.

Another possibility is to include entity definitions in the prolog of your document. Entity definition files [<http://www.w3.org/2003/entities/>] are now maintained by W3C. You can reference definition files with entity definitions you are interested in and then reference imported entities. For example:

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE article [
<!ENTITY % isopub SYSTEM "http://www.w3.org/2003/entities/iso8879/isopub.ent">
%isopub;
]>
<article xmlns="http://docbook.org/ns/docbook" version="5.0">
<title>DocBook V5.0 &ndash; the superb documentation format</title>
...
```

1.3. How to modularize documents?

You can use XInclude [<http://www.w3.org/TR/xinclude/>] for this task. There are available alternative schemas for DocBook V5.0 that contain XInclude elements. This is necessary to make some XML editors happy. Name of these schemas ends with letters “xi”, e.g. `docbookxi.rnc` instead of `docbook.rnc`.

2. Stylesheets

2.1. Will be the current DocBook XSL stylesheets (XSLT 1.0 based implementation) maintained and improved in the future if a new work on a new XSLT 2.0 based implementation started already?

Yes, the current stylesheets (like 1.69.1) will be supported and improved further because they are very widely deployed and work with many existing XSLT processors.

Surely there will be point in a future when all new development will be switched to XSLT 2.0 based implementation only. But this will not happen before all features of the current stylesheets are implemented in the new stylesheets and before there will be more than one usable XSLT 2.0 processor.

Bibliography

[RNCTUT] Clark, James – Cowan, John – MURATA, Makoto: RELAX NG Compact Syntax Tutorial. Working Draft, 26 March 2003. OASIS. <http://relaxng.org/compact-tutorial-20030326.html>

[XMLID] Marsh, Jonathan – Veillard, Daniel – Walsh, Norman: xml:id Version 1.0. W3C Recommendation, 9 September 2005. <http://www.w3.org/TR/xml-id/>

[DB5SPEC] Norman, Walsh: The DocBook Schema. Working Draft 5.0a1, OASIS, 29 June 2005. <http://www.docbook.org/specs/wd-docbook-docbook-5.0a1.html>