

DocBook

A gentle introduction to XML publishing framework
running also under OS/2

Jirka Kosek

University of Economics, Prague

Department of Information and Knowledge Engineering

E-mail: <jirka@kosek.cz>

Web: <http://www.kosek.cz>

Jarda Kačer

E-mail: <jarda@kacer.biz>

Warpstock Europe 2005

November 18–20, 2005

Dresden, Germany

Agenda

- **Part I: Introduction to DocBook**
 - **DocBook and XML**
 - **editing DocBook documents**
 - **processing DocBook documents**
 - **general, platform neutral information**
- **Part II: DocBook Framework for OS/2**
 - **how to install and use DocBook tools in OS/2 environment**
 - **OS/2 specific information**

Introducing myself

- **Jirka Kosek**
 - **freelance consultant, writer and lecturer**
 - **university teacher at University of Economics, Prague – Web and XML classes**
 - **open-source developer—stylesheets for converting DocBook to various output formats**
 - **member of OASIS standardization body—DocBook TC and Entity Resolution TC**

Documentation, DocBook and XML

About DocBook

- **format which is particularly well suited to books and papers about computer hardware and software**
- **can store not only a documentation, but also general books and articles, web-sites, slides, thesis, ...**
- **SGML/XML based**
- **open and platform neutral**
- **stable—DocBook is used and developed since 1991**
- **standardized—DocBook development is done under OASIS**

OASIS takes care about many other open formats, including OpenDocument

- **DocBook users: software vendors (Novell, Sun, HP), publishing houses (O'Reilly), open-source projects (Linux, FreeBSD, PHP, RedHat, ...)**

DocBook Benefits

- **single-source publishing—one source document, many target output formats**
 - **print output (PostScript, PDF, RTF)**
 - **Web output (HTML/XHTML, single page, chunked output)**
 - **help systems (HTMLHelp, info, man, JavaHelp)**
- **support fast software development cycle**
 - **process of generating various output formats is fully automated**
- **modular documents—content reuse**
- **conditional documents**
- **authors no longer have to worry about presentation**

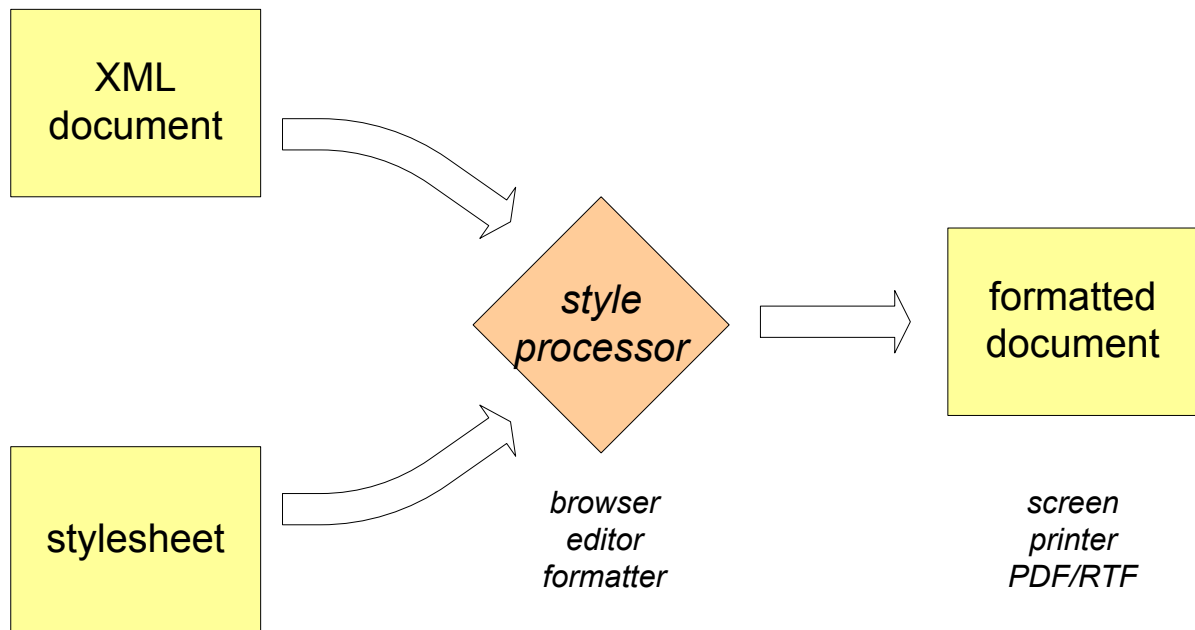
- **simple markup language**

`<para>You can use <application>Emacs</application> for editing your text files.</para>`

- **markup creates documents with a rich information content**
 - **semantics is important**
 - **presentation (formatting) is not important at this stage**
- **XML schema defines set of elements that can be used to markup document content**
 - **various syntaxes for XML schemas: DTD, W3C XML Schema, RELAX NG**

Formatting XML Documents

- separation of content and style
- we can use stylesheet language to describe styling of document
- XML document + style \Rightarrow formatted document



- advantages of stylesheet approach
 - one stylesheet can be used to process many documents (uniform look for a set of documents)
 - one document can be processed by several different stylesheets (multiple output formats)

Overview of Stylesheet Languages

- **CSS (Cascading Style Sheets)**
 - too limited to cover complexity of DocBook
 - are used to define visual appearance of DocBook in visual XML editors
 - W3C recommendation
- **XSL (Extensible Stylesheet Language)**
 - **XSLT** – describes transformation from XML document to XML, HTML or plain text
 - **XSL FO (formatting objects)** – “typographical quality” formatting for printed media (PDF, PostScript, RTF)
 - W3C recommendation
- **DSSSL (Document Style Semantics and Specification Language)**
 - originally developed for SGML, can be used with XML as well
 - transformation to HTML, print output
 - ISO standard

DocBook and XML

- **technically, DocBook is just an XML schema which defines set of elements and attributes allowed in DocBook documents**
- **writing DocBook content is similar to HTML authoring—you just have to learn new element names**
- **there are ready to use DSSSL and XSL stylesheets for DocBook**
- **there are many generic XML tools that can be used to edit, store, manage, transform, query and process DocBook content**

DocBook in 15 Minutes

Sample Document

- there are around 400 elements in DocBook
- you do not have to know all of them to use DocBook

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE book PUBLIC "-//OASIS//DTD DocBook XML ►
V4.4//EN" ►
"http://www.oasis-open.org/docbook/xml/4.4/docbookx.dtd">
<book lang="en">
  <title>My first book written in DocBook</title>

  <chapter>
    <title>Title of the first chapter</title>
    <para>This is a text of paragraph.</para>
    <para>Yet another paragraph.</para>
  </chapter>

  <chapter>
    <title>Title of the second chapter</title>
    <para>Text of paragraph...</para>
    <para>Text of paragraph...</para>
  </chapter>
</book>
```

Structuring Your Content

- **book sets**
- **books or articles—the most used structures**
- **book can have multiple parts**
- **book or part can be subsequently divided to chapters, appendixes, prefaces, table of contents, index, bibliography, reference pages, ...**
- **sections and nested sections**
- **you can assign various metadata to each structural element—author, title, copyright, ...**

Block Elements

- **lists—unnumbered, numbered, glossary, list of term definitions, callouts**
- **admonitions—warning, note, tip, ...**
- **program listings, screen shots, command and function synopsis**
- **tables, examples, equations, procedures, FAQs**
- **images—you can have multiple image variants to satisfy differences between various output formats (raster, vector, ascii-art)**

Paragraphs

- text must be always enclosed in some element, usually paragraph (it is not possible to have free standing text as in HTML)
- two kinds of paragraphs
 - `para`—other block elements can be nested inside `para`
 - `simplpara`—can contain only text or inline elements

source `<para>Sample paragraph. Can be very, very, ... long.</para>`

output Sample paragraph. Can be very, very, ... long.

Lists

- text inside list items must be also enclosed in paragraph
- lists can be nested

source

```
<itemizedlist>
  <listitem>
    <para>apples</para>
  </listitem>
  <listitem>
    <para>pears</para>
  </listitem>
</itemizedlist>

<orderedlist>
  <listitem>
    <para>Reichenberg</para>
  </listitem>
  <listitem>
    <para>Dresden</para>
  </listitem>
</orderedlist>
```

output

- apples
- pears

1. Reichenberg
2. Dresden

Formal Objects

- elements that have title and are usually numbered
- figure
- table
- example
- equation

source

```
<table>
  <title>Sample table</title>
  <tgroup cols="2">
    <thead>
      <row>
        <entry>Entity</entry>
        <entry>Description</entry>
      </row>
    </thead>
    <tbody>
      <row>
        <entry>&amp;&&</entry>
        <entry>ampersand</entry>
      </row>
      <row>
        <entry>&lt;</entry>
        <entry>less than symbol</entry>
      </row>
    </tbody>
  </tgroup>
</table>
```

Formal Objects (Continued)

output **Table 1. Sample table**

Entity	Description
&amp;	ampersand
&lt;	less than symbol

Images

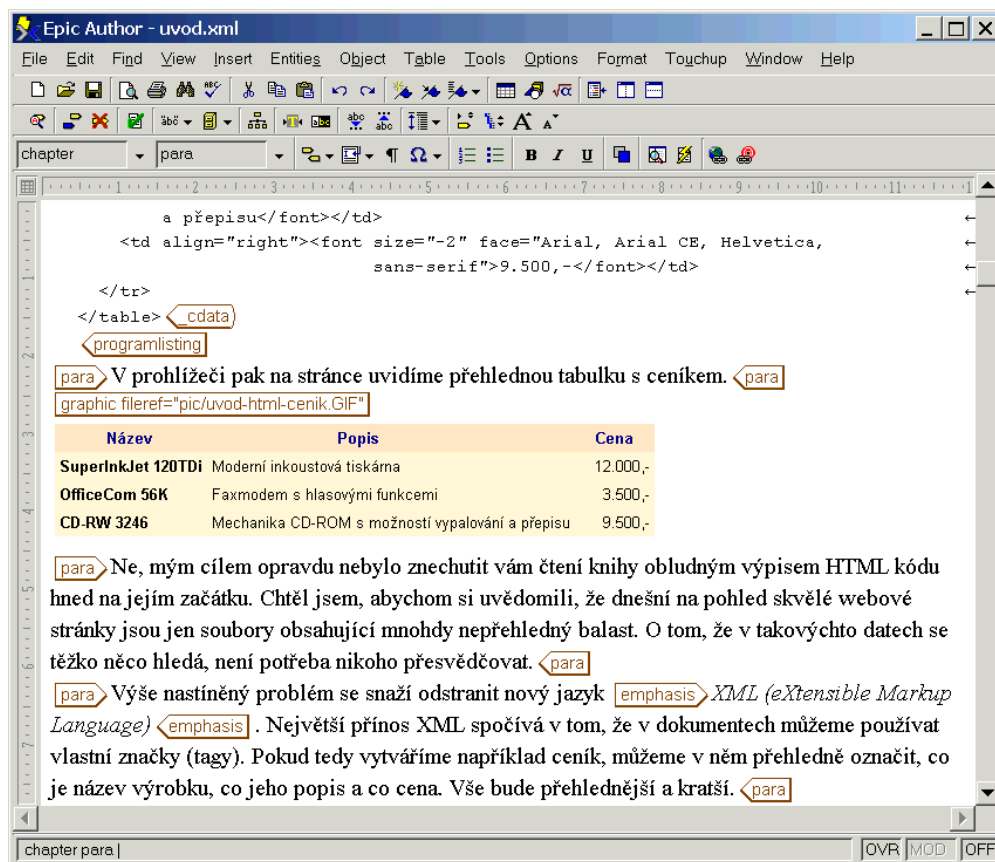
- each image can be provided in multiple formats—e.g. raster image for Web and vector variant for printing
- the most suitable image variant is automatically selected during processing
- **block** (`mediaobject`) and **inline** (`inlinemediaobject`) images
- image with title is created via `figure` formal object
- image data are referenced and modified using `imagedata` element

```
source    <figure id="editor">
            <title>Epic editor in action</title>
            <mediaobject>
                <imageobject>
                    <imagedata fileref="image.svg"/>
                </imageobject>
                <imageobject>
                    <imagedata fileref="image.png"
                               width="100%"/>
                </imageobject>
                <textobject>
                    <phrase>Screenshot of DocBook ►
document edited inside Epic ►
editor.</phrase>
                </textobject>
            </mediaobject>
        </figure>
```

Images (Continued)

output

Figure 1. Epic editor in action



Links

- **xref**—cross-reference with an automatically generated content
- **link** – cross-reference with a manually created content

source `<para>There is a screenshot of
the leading XML editor
on <xref linkend="editor"/>.</para>`

```
<para>There is a screenshot of the leading
XML editor on
<link ►
linkend="editor">figure</link>.</para>
```

output **There is a screenshot of the leading XML editor on Figure 1, “Epic editor in action”.**

There is a screenshot of the leading XML editor on figure.

External Links

- **ulink**—reference to any URL
- **olink**—links to other DocBook documents
 - solves problems like “PDF file should reference another PDF file, while HTML version should reference other HTML pages”

source `<para>You can find more information about
this topic <ulink ►
url="http://www.kosek.cz">on
my home page</ulink>.</para>`

output You can find more information about this
topic on my home page¹.

¹ <http://www.kosek.cz>

Inline Elements

- capture meaning of short text phrases (words)
- general—computer output, user input, replaceable value, XML/HTML tag, ...
- user interface—buttons, icons, menus, keys, hot keys, mouse buttons, ...
- programming—constant, variable, class, function, interface, database, error code, error message, struct, token, property, return value, ...
- operating system—application, command, environment variable, filename, message, option, parameter, ...

source `<para>You must set <envar>APPHOME</envar>`
environment variable in order to get
application running. On unix system
you set the variable by issuing command
`<command>APPHOME=/usr/local/app; export`
`APPHOME</command>`. In Windows, go to
`<application>Control panel</application>`
and then inside ►
`<guimenuitem>System</guimenuitem>`
property card press
`<guibutton>Environment</guibutton>`.
User names are stored inside
`<filename>/etc/passwd</filename>` file.
My email address is
`<email>jirka@kosek.cz</email>`. You can
print to stdout by ►
`<function>printf</function>`

Inline Elements (Continued)

```
function. Computer can be shutted down by ►  
pressing  
<keycombo ►  
action="simul"><keycap>Ctrl</keycap>  
<keycap>Alt</keycap><keycap>Del</keycap></keycombo>.</para>
```

output **You must set APPHOME environment variable in order to get application running. On unix system you set the variable by issuing command APPHOME=/usr/local/app; export APPHOME. In Windows, go to Control panel and then inside System property card press Environment. User names are stored inside /etc/passwd file. My email address is <jirka@kosek.cz>. You can print to stdout by printf function. Computer can be shutted down by pressing Ctrl+Alt+Del.**

DocBook Friendly Editors

Editors

- **you can use any plain text editor**
- **it is better to use editors which are specialized for XML, or even for DocBook editing, because you can get:**
 - **syntax highlighting**
 - **schema driven markup completion**
 - **validation**
 - **WYSIWYG-like editing**
 - **preinstalled DocBook schema**
 - **preinstalled DocBook stylesheets**

Emacs

- <http://www.gnu.org/software/emacs/>
- two specialized XML modes PSGML² and nXML³
- syntax highlighting
- schema driven markup completion
- not WYSIWYG, you work directly with markup (not suitable for “ordinary” users)
- the most powerful editor in the world
- it takes some time to get familiar with not so conventional UI
- runs on almost any platform
- free, GPL license

² <http://sourceforge.net/project/psgml/>

³ <http://www.thaiopensource.com/nxml-mode/>

- <http://www.jedit.org/>
- XML plugin⁴ turns jEdit into a quite powerful XML editor
- syntax highlighting
- schema driven markup completion (though not 100% accurate)
- can validate document
- not WYSIWYG, you work directly with markup (not suitable for “ordinary” users)
- written in Java—runs everywhere
- free, GPL license

⁴ <http://plugins.jedit.org/>

XML Mind XML Editor

- <http://www.xmlmind.com/xmleditor/>
- two versions: Standard (free) and Professional (paid)
- free Standard version can be used even for commercial purposes
- visual, WYSIWYG-like editing of XML documents
- visual table editor
- preconfigured with DocBook schemas and stylesheets
- schema driven markup completion
- can validate document
- very good support for modular documents
- written in Java—runs everywhere

Another Commercial Editors

- **oXygen⁵**
 - **syntax highlighting and schema driven markup completion**
 - **comes with preconfigured DocBook support**
 - **written in Java**
- **XMLSpy⁶**
 - **both WYSIWYG and source code view**
 - **tag name completion is not 100% perfect**
 - **Windows only**
- **Epic⁷**
 - **leading WYSIWYG-like XML editor**
 - **preconfigured with DocBook support**
 - **integrated TeX-based print engine**
 - **Windows, Solaris**
- **XMetaL⁸**
 - **very user friendly WYSIWYG-like XML editor**
 - **no out of the box DocBook support**
 - **Windows only**

⁵ <http://www.oxygenxml.com/>

⁶ <http://www.altova.com/>

⁷ http://www.arbortext.com/html/epic_editor_overview.html

⁸ <http://xmetal.com/>

Formatting and Converting

Free Formatting Tools

- **DocBook stylesheets**⁹
 - two independent implementations—DSSSL and XSL
 - originally created by Norman Walsh
 - now maintained by a group of several people
- **DSSSL stylesheets**
 - we need DSSSL processor to use them—Jade¹⁰ or OpenJade¹¹
 - supported output formats: HTML, chunked HTML, RTF, TeX (and thus PDF and PostScript), FrameMaker
- **XSL stylesheets**
 - we need arbitrary XSLT processor (e.g. Saxon, xsltproc, Xalan)
 - supported output formats: HTML, XHTML, chunked HTML, JavaHelp, HTMLHelp, FO, Eclipse Help, man, WordML
 - printed output (PDF) is generated from FO using rendering engine like XEP, XSL Formatter or FOP

⁹ <http://docbook.sourceforge.net/>

¹⁰ <http://www.jclark.com/jade/>

¹¹ <http://openjade.sourceforge.net/>

DocBook XSL Stylesheets

- the most commonly used tool for converting DocBook to various target formats
- easy to customize
 - using parameters
 - overriding XSLT templates
- good internationalization—support for more than 50 languages
- advanced functionalities
 - profiling (conditional documents)
 - internationalized indexing

Another Free Formatting Tools

- **docbook2X¹²**
 - **conversion to man and info**
- **<SGML>&tools;¹³**
 - **shell scripts for easy invocation of DSSSL based tools**
- **DB2LaTeX¹⁴**
 - **XSLT stylesheets for converting DocBook to LaTeX**
- **DBLaTeX¹⁵**
 - **XSLT stylesheets for converting DocBook to LaTeX and ConTeXt**

¹² <http://docbook2x.sourceforge.net/>

¹³ <http://sgmltools-lite.sourceforge.net/>

¹⁴ <http://db2latex.sourceforge.net>

¹⁵ <http://dblatex.sourceforge.net>

DocBook Future

DocBook Future

- the first betaversion of DocBook V5.0 was released three weeks ago
 - improved validation thanks to RELAX NG + Schematron based schemas
 - DTD and W3C XML Schema are also provided for legacy tools
 - simplified several content models
 - all DocBook elements are now in a namespace
 - every element can work as a link (XLink)
- new XSLT 2.0 based version of the DocBook XSL stylesheets is under development
 - no need for extensions, pure XSLT 2.0 code
 - advanced things like profiling or internationalized indexing are enabled by default

Resources

DocBook Community

- **discussion about tools issues –**
`<docbook-apps@lists.oasis-open.org>` (archive¹⁶)
- **discussion about DocBook schema issues –**
`<docbook@lists.oasis-open.org>` (archive¹⁷)
- **WiKi—<http://wiki.docbook.org/>**
- **report bugs¹⁸ (it improves software!)**
- **OASIS DocBook TC**
 - **OASIS: The Organization for the Advancement of Structured Information Standards**

A non-profit, international consortium that creates interoperable industry specifications based on public standards such as XML and SGML. OASIS members include organizations and individuals who provide, use and specialize in implementing the technologies that make these standards work in practice.
 - **technical committee which further develops and maintains DocBook schemas**
 - **“adds new elements and attributes”**
 - **monthly teleconferences**
- **DocBook Developers**
 - **informal group of XSL and DSSSL stylesheets developers**

¹⁶ <http://lists.oasis-open.org/archives/docbook-apps/>

¹⁷ <http://lists.oasis-open.org/archives/docbook/>

¹⁸ https://sourceforge.net/tracker/?group_id=21935&atid=373747

Documentation

- **Official DocBook home**¹⁹
- **DocBook reference documentation and schema repository**²⁰
- **XSL and DSSSL stylesheets**²¹
- **DocBook XSL: The Complete Guide**²² – everything you need to know about using and customizing XSL stylesheets
- **DocBook V5.0: The Transition Guide**²³—how to migrate from DocBook V4.x to V5.0
- **These slides:**
<http://www.kosek.cz/xml/2005warpstockeu/>

¹⁹ <http://www.oasis-open.org/docbook/>

²⁰ <http://www.docbook.org>

²¹ <http://docbook.sourceforge.net>

²² <http://www.sagehill.net/docbookxsl/>

²³ <http://docbook.org/docs/howto/>