

DocBook

Textverarbeitung mit XML

Peter Eisentraut

5.3.2005

Persönlich

- PostgreSQL-Entwickler
- Maintainer PostgreSQL-Dokumentation
- Debian-Entwickler
- Debian XML/SGML Group
- Maintainer DocBook DSSSL Stylesheets
- Buchautor

Was ist DocBook?

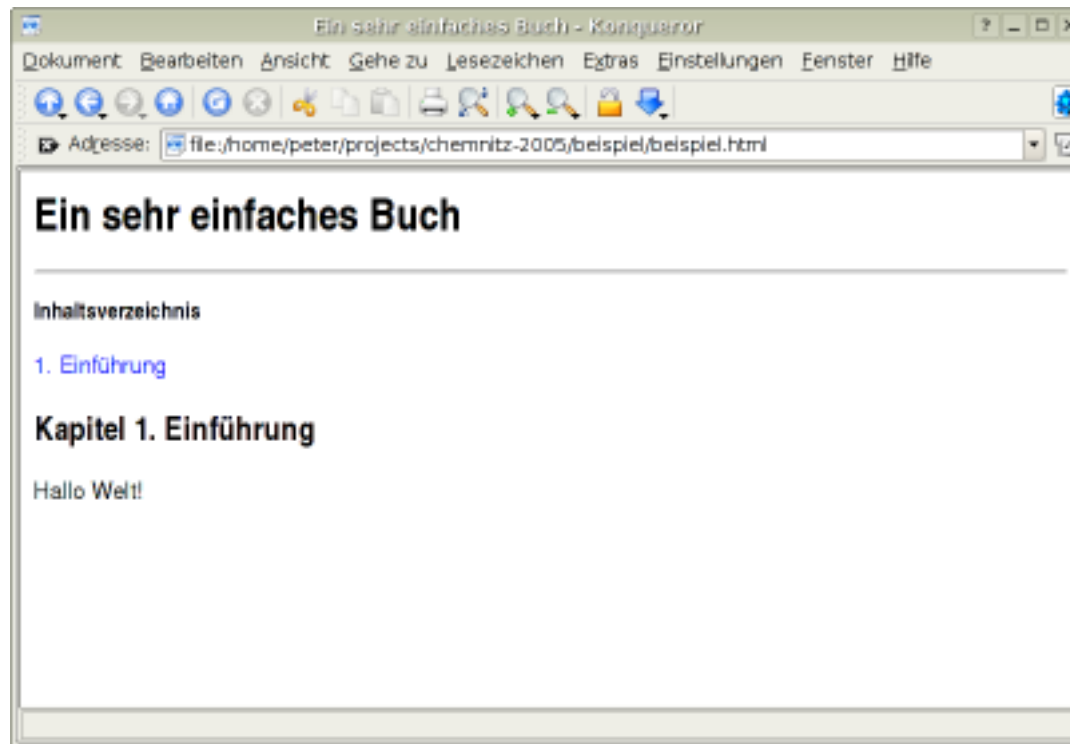
- XML-Vokabular für technische Dokumente
- Stylesheets und viele Tools
- OASIS-Standard

Ein einfaches Beispiel

```
<?xml version="1.0"?>
<!DOCTYPE book
  PUBLIC "-//OASIS//DTD DocBook XML V4.3//EN"
  "http://www.oasis-open.org/xml/4.3/docbookx.dtd">
<book lang="de">
  <title>Ein sehr einfaches Buch</title>
  <chapter>
    <title>Einf&uuml;hrung</title>
    <para>Hallo Welt!</para>
  </chapter>
</book>
```

Verarbeitungsbeispiel

```
$ xsltproc /usr/share/xml/docbook/stylesheet/nwalsh/html/docbook.xsl \
  beispiel.xml > beispiel.html
```



Vorteile von DocBook

- Inhalt und Format getrennt
- mehrere Ausgabeformate
- XML als standardisiertes Format
- erweiterbar, anpassbar
- läuft auf vielen Plattformen

Ausgabeformate

- HTML, XHTML
- PDF, PostScript
- HTML Help, JavaHelp
- RTF
- Manpages
- Plain Text
-

Nachteile von DocBook

- kein WYSIWYG
- kein DTP-Ersatz
- nicht einfach
 - viele Tags (HTML: 90, DocBook: 400)
 - strikte Struktur
 - viele kooperierende Tools
- perfekte Druckergebnisse schwierig

Einsatzbereiche

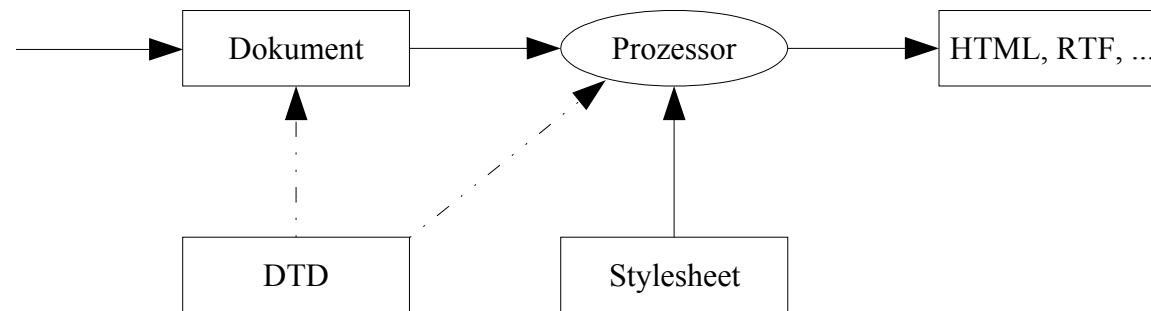
- wiederholte Releases
- verteilte Autorenschaft
- große, komplexe Dokumente
- mehrere Ausgabeformate
- automatische Verarbeitung

Aktuelle Anwender

GNOME, KDE, FreeBSD, Linux-Kernel, LDP, PostgreSQL, Debian, OpenLDAP, PHP, SuSE, Red Hat, Sun, O'Reilly, . . .

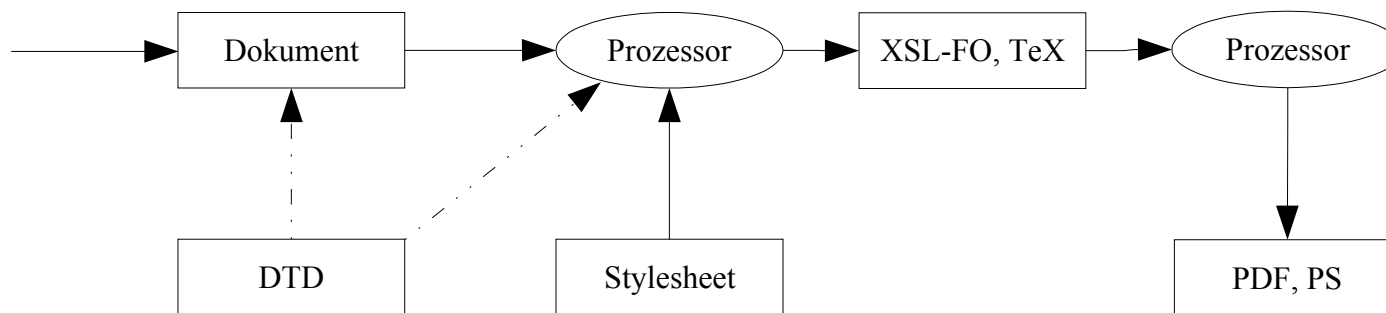
Verarbeitung

Umwandlung von DocBook in Zielformat



Verarbeitung

Umwandlung von DocBook ber Zwischenformat in Zielformat



Die DocBook DTDs

Document Type Definition — legt die Dokumentstruktur fest

- für XML und SGML
- Aktuell Version 4.4
- Veröffentlicht von OASIS

Zukünftig (Version 5):

- Relax NG Schemas
- XML Schemas (experimentell)

DTD-Beispiel

```
<!ELEMENT Chapter
((DocInfo?, Title, Subtitle?, TitleAbbrev?), ToCchap?,
(((CalloutList | GlossList | ItemizedList | OrderedList | SegmentedList |
SimpleList | VariableList | Caution | Important | Note | Tip | Warning |
LiteralLayout | ProgramListing | ProgramListingCO | Screen | ScreenCO |
ScreenShot | Synopsis | CmdSynopsis | FuncSynopsis | FormalPara | Para |
SimPara | Address | BlockQuote | Graphic | GraphicCO | MediaObject |
MediaObjectCO | InformalEquation | InformalExample | InformalFigure |
InformalTable | Equation | Example | Figure | Table | MsgSet | Procedure |
Sidebar | QandASet | Anchor | BridgeHead | Comment | Highlights |
Abstract | AuthorBlurb | Epigraph | IndexTerm)+,
(Sect1* |
(RefEntry)* |
SimpleSect* | Section*))|
(Sect1+ |
(RefEntry)+ |
SimpleSect+ | Section+)),
(Index | Glossary | Bibliography)*)
>
```

Stylesheets

- XSLT-Stylesheets
 - W3C-Standard
 - nur für XML DTD
 - Ausgabe: HTML, XHTML, HTML Help, XSL-FO, . . .
- DSSSL-Stylesheets
 - ISO-Standard
 - für XML und SGML DTDs
 - Ausgabe: HTML, T_EX, RTF

XSLT-Beispiel

DocBook: Schreiben Sie an `<email>user@domain.com</email>`.

```
<xsl:template match="email">
  <xsl:call-template name="inline.monoseq">
    <xsl:with-param name="content">
      <xsl:text>&lt;</xsl:text>
      <a>
        <xsl:attribute name="href">mailto:<xsl:value-of select="."/></xsl:attribute>
        <xsl:apply-templates/>
      </a>
      <xsl:text>&gt;</xsl:text>
    </xsl:with-param>
  </xsl:call-template>
</xsl:template>
```

HTML: Schreiben Sie an `<tt><user@domain.com></tt>`.

DSSSL-Beispiel

DocBook: Schreiben Sie an `<email>user@domain.com</email>`.

```
(element email
  ($code-seq$
    (make sequence
      (literal "&#60;")
      (make element gi: "A"
        attributes:
          (list (list "HREF"
                    (string-append "mailto:"
                                   (data (current-node))))))
        (process-children))
      (literal "&#62;")))))
```

HTML: Schreiben Sie an `<tt><user@domain.com></tt>`.

XSLT-Prozessoren

- xsltproc
- Saxon
- Xalan
- . . .

XSL-FO-Prozessoren

Frei:

- FOP
- Passive $\text{T}_{\text{E}}\text{X}$

Nicht frei:

- Antenna House XSL Formatter
- RenderX XEP
- Unicorn Formatting Objects

DSSSL-Toolchain

- Umwandlung DSSSL: Jade, OpenJade
- Umwandlung $\text{T}_{\text{E}}\text{X}$ in PDF: Jade $\text{T}_{\text{E}}\text{X}$

Schreibwerkzeuge

- Texteditor
- Editor mit SGML/XML-Modus
- XML-Editor

Beispiele: Inline-Markup

`<para>Auf <trademark class="registered">UNIX</trademark>-Systemen steuert die Datei <filename>/etc/crontab</filename> die von <command>cron</command> ausgeführten Befehle.</para>`

`<para><symbol>UINTMAX</symbol> ist gleich 2<superscript>32</superscript>-1.</para>`

Beispiele: Listen

```
<itemizedlist>  
  <listitem>  
    <para>Erster Punkt</para>  
  </listitem>  
  <listitem>  
    <para>Zweiter Punkt</para>  
  </listitem>  
  <listitem>  
    <para>Dritter Punkt</para>  
  </listitem>  
</itemizedlist>
```

Beispiele: Formale Objekte

Abbildungen, Beispiele, Gleichungen, Tabellen

- automatisch nummeriert
- automatisches Abbildungsverzeichnis usw.

Beispiele: Code

<para>Dies ist ein sehr einfaches Shell-Skript:

```
<programlisting>
```

```
#!/bin/sh
```

```
exit 0
```

```
</programlisting>
```

```
</para>
```

Beispiele: Querverweise

```
<chapter id="maus">  
<title>Verwendung einer Maus</title>  
...  
Siehe <xref linkend="maus"/>.
```

HTML-Ausgabe:

```
Siehe <a href="#maus">Kapitel 3, Verwendung einer Maus</a>.
```

Beispiele: Grafiken

Unterstützte Formate hängen von der Toolchain ab.

```
<mediaobject>
  <imageobject>
    <imagedata fileref="figures/eiffeltower.eps" format="eps">
  </imageobject>
  <imageobject>
    <imagedata fileref="figures/eiffeltower.gif" format="gif">
  </imageobject>
  <textobject>
    <phrase>The Eiffel Tower</phrase>
  </textobject>
</mediaobject>
```

Customization

DocBook DTD anpassbar:

- Elemente hinzufügen
- Elemente entfernen

DocBook Stylesheets anpassbar:

- neue Formatierungsregeln

HTML-Ausgabe anpassbar:

- CSS

Alternative Anwendungen

- Simplified DocBook DTD
- Module: EBNF, HTML Forms, MathML, SVG
- Website DTD
- Slides DTD
- angepasste Rechtschreibprüfung
- Gettext-Integration

Zusammenfassung

DocBook . . .

- ist eine XML DTD für technische Dokumente
- unterstützt viele Ausgabeformate
- ist geeignet für wiederholte und verteilte Bearbeitung
- ist nicht immer ganz einfach

Weitere Informationen:

- <http://docbook.sourceforge.net/>