

Publishing mit XML

Jochen Hein

Copyright © 2004 Jochen Hein

Inhalt

- **Publishing mit XML: Warum eigentlich?**
- **DocBook als XML-Applikation**
- **Workflow und Tools**
- **Erfahrungen beim Publishing: Pearson Education**
- **Fazit**

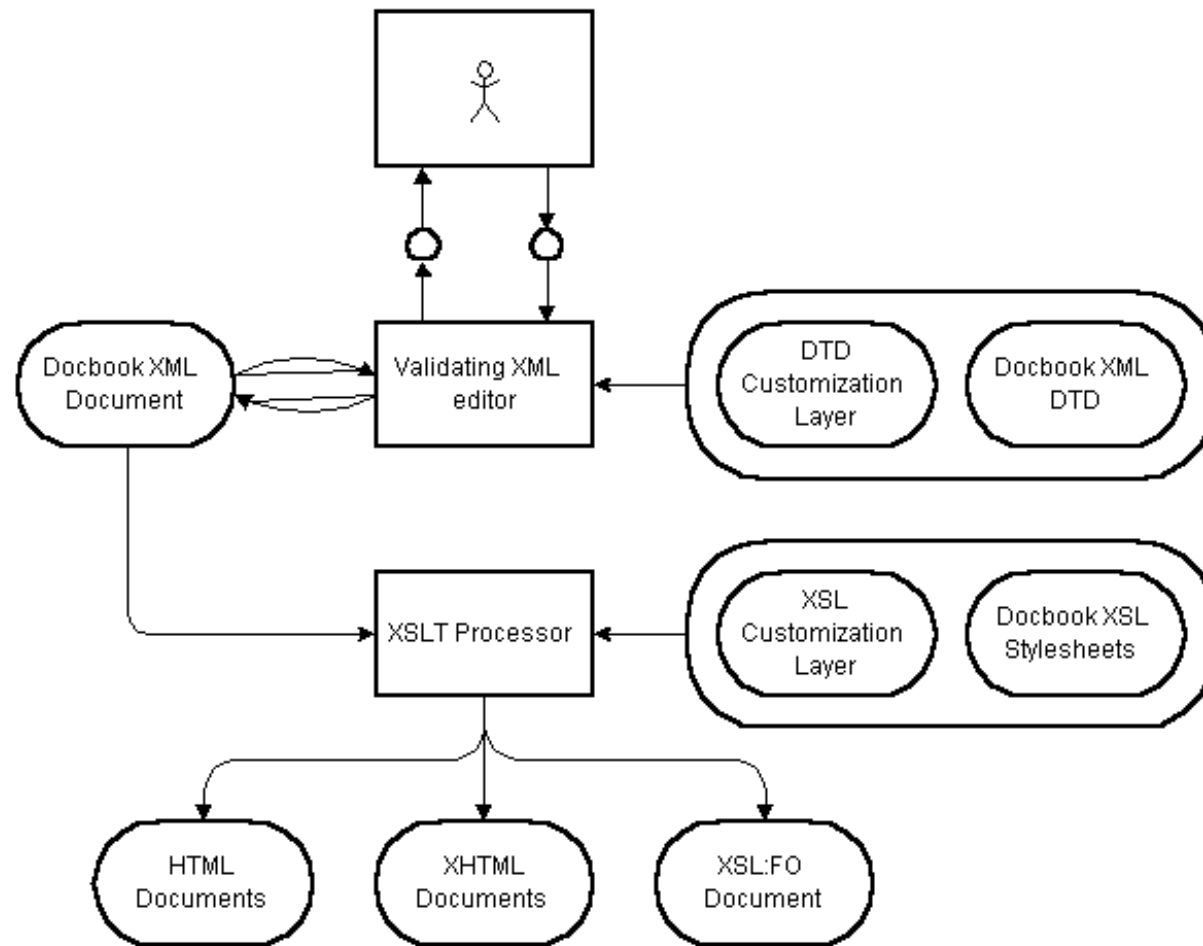
Warum Publishing mit XML?

- **Standardformat (W3C und Applikationen bzw. DTDs)**
- **Maschinell konvertierbar in verschiedene Formate**
- **Plain ASCII: Editierbar, Versionsverwaltung, diff und Co.**
- **Tools und Stylesheets verfügbar und stabil**
- **Kritische Masse durch große Open Source Projekte**

DocBook als XML-Anwendung

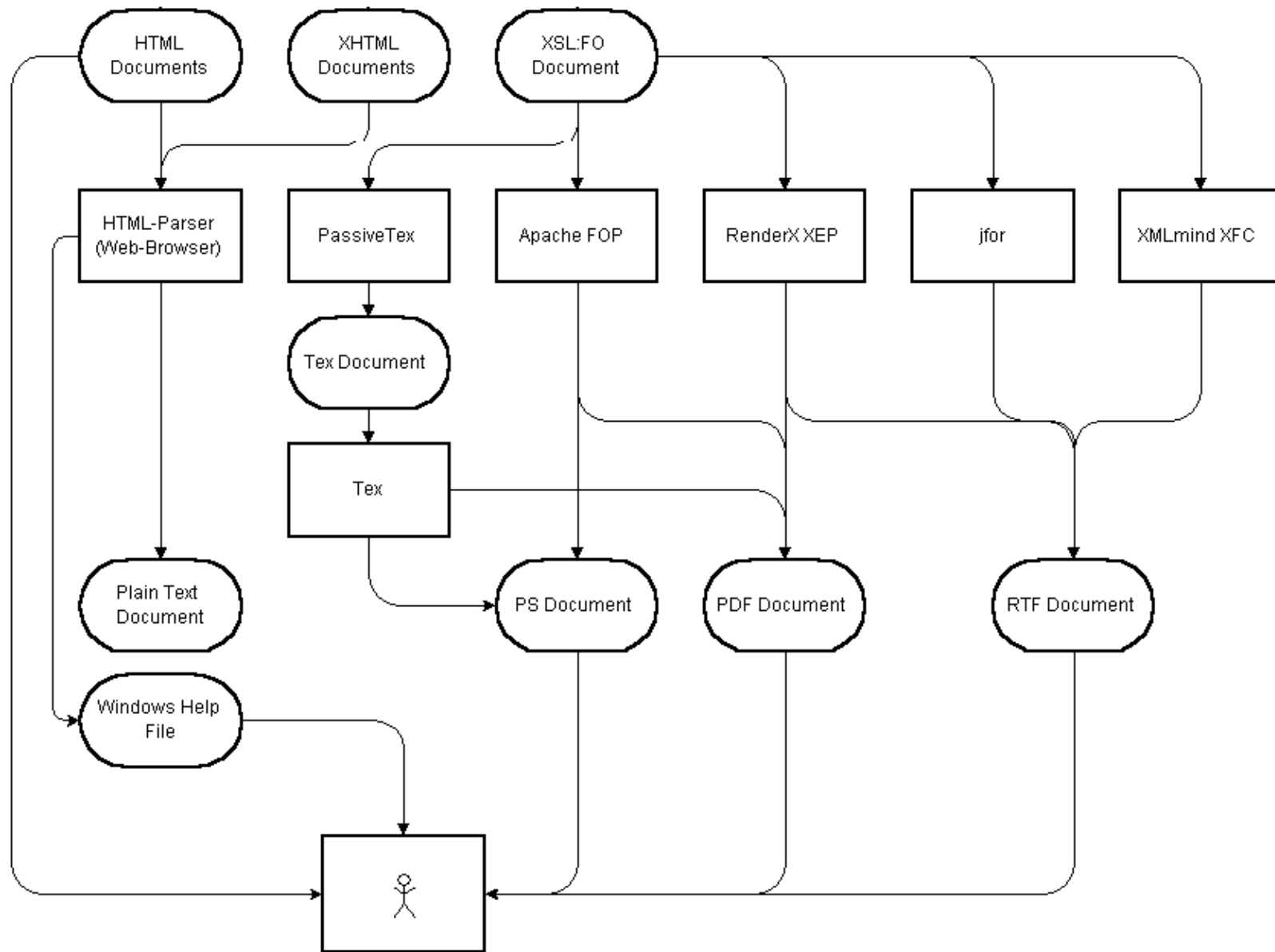
- **Stabile und ausgereifte DTD für Publikationen im Computer-Umfeld**
- **Strukturen für Artikel und Bücher, Erweiterungslayer für Präsentationen und Webseiten**
- **Verfügbarkeit von Stylesheets für HTML und Print, XSL und DSSSL**
- **Dokumentierte Kompatibilität für die nächste Major Release**
- **Verwendung durch kommerzielle Unternehmen und Open Source Projekte**
- **Guter Support auf der docbook-apps-Liste**

Workflow und Tools (1)



Aus dem DocBook Tutorial [<http://trieloff.net/docbook/tutorial/>]

Workflow und Tools (2)



Verlagseigene DTD (Pearson-DTD)

- **Eher visuelles Markup**
- **Original-Manuskript in LaTeX mit logischem Markup**
- **Konvertierung nach DocBook halb-automatisch, kein Informationsverlust**
- **Zum Satz Konvertierung mittels XSL in die Pearson-DTD mit 850 Zeilen Stylesheet**
- **Quick&Dirty: Korrekturlesen des Dokumentes im Pearson-Format mit angepassten DocBook-Stylesheets**
- **Feedback: belichtetes XML-Dokument zur Korrektur der Stylesheets**

Vergleich DocBook und Pearson: Anhang

<pre><appendix> <title>Titel</title> <para>...</para> </appendix> <appendix> <title>Titel 2</title> <para>...</para> </appendix></pre>	<pre><appendix> <chapter> <title>Titel</title> <p>...</p> </chapter> <chapter> <title>Titel 2</title> <p>...</p> </chapter> </appendix></pre>
--	---

Konvertierung DocBook nach Pearson: Anhang

```
<xsl:template match="appendix[1]">
  <appendix>
    <chapter>
      <xsl:call-template name="transform.id.attribute"/>
      <xsl:apply-templates/>
    </chapter>
    <xsl:for-each select="following-sibling::appendix">
      <chapter>
        <xsl:call-template name="transform.id.attribute"/>
        <xsl:apply-templates/>
      </chapter>
    </xsl:for-each>
  </appendix>
</xsl:template>

<xsl:template match="appendix[position() != 1]">
</xsl:template>
```

Konvertierung DocBook nach Pearson: title, para

```
<xsl:template match="para">  
  <p><xsl:apply-templates/></p>  
</xsl:template>
```

```
<xsl:template match="title">  
  <title><xsl:apply-templates/></title>  
</xsl:template>
```

Konvertierung LaTeX nach XML

Erster Wurf mit ltx2html mit eigener Steuerdatei

Relativ mechanische, aber manuelle Nacharbeit notwendig

Problemfälle:

- **Index-Einträge: verschiedene Processing-Expectations**
- **Tabellen nicht maschinell konvertierbar**
- **Umlaute sind problematisch**
- **Unbekannte LaTeX-Befehle werden ignoriert - manuelle Durchsicht des Ergebnisses erforderlich**

Erfahrungen: XML-Dokumente

- **Gewohnte Editor-Umgebung: Emacs mit PSGML**
- **Einsatz von CVS für die Arbeit auf verschiedenen Rechnern**
- **Regelmäßige Validierung mittels xmllint und Co.**
- **Erste Korrekturfahnen mittels DocBook Stylesheets**
- **Erste, Proof-of-Concept Konvertierung in die Pearson-DTD**
- **Validierung gemäß der Pearson-DTD um formale Fehler aufzudecken**
- **1700 Zeilen XSL-Customizing für Druck, meist "kreatives" Cut&Paste der DocBook-Styles**

Erfahrungen: Ausgabe-Formate

- **Online-Korrektur: HTML-Format, gute Styles, schnelle Konvertierung**
- **Print-Korrektur: Pearson-Format und dann PDF: brauchbare Styles und Tools, langsame Konvertierung**
- **Andere Formate wie RTF via SGML/DSSSL-Tools: brauchbare Styles**

Fazit

- **Gute Tools zum Editieren und Verwalten der Dokumente**
- **Einfache Änderung der Styles an Sonderwünsche**
- **Die Wahl von DocBook hat sich bewährt, auch wenn die Ausgabe im PDF-Format Schwächen hat**
- **Persönliche würde ich diese Wahl immer wieder treffen**
- **Ob diese Wahl in allen Umgebungen sinnvoll ist darf bezweifelt werden**

Fragen...

... und Antworten

Die Folien, Styles und das Paper gibt's unter <http://www.jochen.org/~jochen/vortraege.html>