

Office meets XML

XML-Schnittstellen von Bürosoftware (konkret: OpenOffice)

W. Riedel, TU Chemnitz, URZ

<mailto:w.riedel@hrz.tu-chemnitz.de>

 531 1422

Workshop, 19.-22.4.2004, Löbsal

Was erwartet uns?

1	Das Aufgebot	3
2	Die Verlobten: OpenOffice und DocBook	5
3	Polterabend	8
4	Die Hochzeit	9
5	Probleme des Alltags	10
6	Nebenbuhler	14

1 Das Aufgebot: Worum geht es eigentlich?

Motivation:

Artikel im Linux-Magazin 11/03: „Schreib-Hilfe“, S. 92ff:

OpenOffice kann DocBook verarbeiten und erzeugen

XML-Applikation „DocBook“ schon lange bekannt und in Gebrauch:

- Linux-, FreeBSD-Entwickler
- KDE-, Gnome-Entwickler (z.B.
`/usr/share/doc/HTML/de/*/* .docbook`)
- in Firmen (Sun, HP)

(meine) Vision:

- DocBook als standardisiertes Datenformat zum Austausch zwischen beliebigen Anwendungen
- Dokumenteneingabe mit wysiwyg-orientierten Werkzeugen
- Formatierung mit „richtiger“ Satzsoftware
- beliebige semantik-basierte Weiterverarbeitung der Informationen

2 Die Verlobten: OpenOffice und DocBook

Der Bräutigam: DocBook

- <http://docbook.org/>, <http://www.oasis-open.org/docbook/>
- DocBook ist SGML- als auch XML-Applikation
- „System zum Schreiben strukturierter Dokumente, besonders geeignet für Bücher und Fachtexte über Computerhard- und -software“
- aktueller Stand: 4.3RC3
- definiert durch DTD, etwa 350 Tags

⇒ keine praktikable Schnittstelle für den Mensch

Modifikation: „Simplified DocBook“

- <http://docbook.org/specs/wd-docbook-simple-1.1b1.html>
- aktueller Stand: 1.1b1
- etwa 100 Tags, also immer noch ziemlich „geschwätzig“

Ein Beispieldokument

⇒ eigentlich immer noch (nur) maschineninternes Format

Die Braut: OpenOffice

- OO-Format ist XML!
- aber nach dem unzip: 😊
- aber: XML \neq DocBook

⇒ die Braut ziert sich noch ziemlich ...

3 Polterabend

Konverter: OOo2sDBK

- www.chez.com/ebellot/ooo2sdbk
- Autor: Eric Bellot
- Doku nur französisch
- benutzt Saxon + Python
- Konvertierung fehlerhaft
- Einbahnstraße!

4 Die Hochzeit

- spezielle Konfiguration notwendig:
Menü EXTRAS – XML-FILTEREINSTELLUNGEN
- dazu sind Files von www.openoffice.org notwendig (bei uns als rpm verteilt)
- jetzt kann man SPEICHERN UNTER im DocBook-Format
- Öffnen vorhandener Dokumente analog, OO wendet passende XSL an

Der „Beweis“

5 Probleme des Alltags

- voriges Beispiel: Anzeige des Dokumentkopfs problematisch, ansonsten okay
- wie erfolgt semantik-gerechte Eingabe?
was passiert mit visuellem Markup?
⇒ Dokumentvorlagen
= Umsetzung der DTD und Verknüpfung zu Style
- also Dokument-Vorlage „DocBook (Article)“ öffnen, dann zeigt der Stylist (**Absatzvorlagen**, **Zeichenvorlagen**) mit englischen Bezeichnungen die Elemente, die in DocBook abgebildet werden können

nun kann es losgehen - ein erster Versuch: **OpenOffice**, **Ergebnis**

Ausgabe falsch

- Untertitel fehlt
- `<sect1>` falsch geschachtelt
- keine Liste erzeugt
- kein Bezug auf S-DocBook

Öffnen vorhandener DocBook-Files

- liefert keinen Fehler, Tags werden akzeptiert, aber nicht verarbeitet
- keine Formatierung sichtbar

Prinzipiell

- Kenntnis der DocBook-Tags notwendig
- verschachtelte Tags nur eingeschränkt möglich
Workaround: Arbeit mit Bookmarks, Bereichen, globale Variablen (für ArticleInfo)
- OO kennt nicht alle Tags von SDB, dafür einige aus DB, die nicht zu SDB gehören

Resümee: (noch?) nicht ganz befriedigend

Was kann mit damit eigentlich machen?

- einem Browser anbieten, dazu muss noch eine Style-Deklaration eingefügt werden

```
<?xml-stylesheet href="sdocbook.css" type="text/css"?>
```

Mozilla zeigt „etwas“

- Weiterverarbeitung mittels
 - db2xx (Jade) [SGML-DocBook]
 - xsltproc [XML-DocBook]
 - xmlto
 - docman (DocBook Conversion Toolchain Manager)
 - Saxon, Xalan, FOP, ConT_EXt o.ä.

6 Nebenbuhler

OpenOffice unter WXP

- XML-Filtereinstellungen bereits standardmäßig vorhanden
- Dokumentvorlage muss in PFADE konfiguriert werden
- dann wie unter Linux

MS Office

- ab Office 2003 ist Schnittstelle zu DocBook vorhanden
- muss per Hand konfiguriert werden
- bisher keine praktischen Erfahrungen

AbiWord (GNOME Office)

- Version 2.0.1 (Fedora)
- Abiword-Format (.abw) ist XML
- kann DB ausgeben (SPEICHERN UNTER)
- öffnen unklar
- existieren Plugins für DocBook – nicht installierbar
- „semantik-orientierte Nutzerführung“ nicht erkennbar

Kword (KOffice)

- Version 1.2.1 (Fedora)
- „KWord can save DocBook documents in the SGML version only. It saves texts, lists, pictures, tables and hyperlinks. Only the simplest formatting (bold and italic) are supported. Pictures are saved into a sub-directory.“
- „KWord cannot load any DocBook document.“

LyX

- Version 1.3.3
- es gibt eine Vorlage „Docbook (Article)“
- damit wird die Struktur des **Dokumentenkopfs** vorgegeben
- Hauptmenü liefert verfügbare Tags
- exportieren als Docbook (SGML) **möglich**
- Import nur für „Linuxdoc“

Danke für die Aufmerksamkeit