

Automatische kostenlose PDF-Erzeugung aus einer XML-Datenbasis

Alexander Konzelmann

*Richard Boorberg Verlag
D-70563 Stuttgart, Scharrstr. 2
a.konzelmann@boorberg.de*

Schlagnote: XML, SGML, Java, FOP (formatting objects processor), automatischer Satz, kostenloses PDF, Voraussetzungen, Vorgehensweise

Abstract: XML-strukturierte Datenhaltung eröffnet konkrete Einsparmöglichkeiten für Verlage und die öffentliche Hand. Eine solche bietet sich bei der Erzeugung von Satz für Papierausgaben. XML-Dateien können von einem Java-Programm kostenlos und automatisch in PDF umgewandelt werden.

1. Einführung

XML-strukturierte Daten helfen bei der Lösung gewisser Aufgabenstellungen der juristischen Fachpublikation.

Zuerst zu den Begrifflichkeiten: „XML“ kann synonym zu „SGML“ verstanden werden, weil die SGML-Dateien („Instanzen“), von denen die Rede ist, zwar auf einer SGML-notierten DTD (Document Type Definition) basieren, jedoch immer auch wohlgeformtes XML darstellen. Die meisten Publikationsaufgaben setzen keine XML-notierte DTD voraus. SGML-Instanzen sind durch Umschreiben der Umlaute auf Unicode UTF-8, Ergänzen von Ende-Tags zu leeren Elementen und durch Austausch der DOCTYPE-Zeile durch einen XML-Dateikopf mit Verweis auf ein Stylesheet zu tauglichem Ausgangsmaterial für die erforderlichen Rendering-Systeme zu machen.

Dann zur Darstellung im Netz: Es genügt ein XSL-Stylesheet, welches aus den hierarchischen Gliederungsebenen ein Übersichtsdokument mit Links zu den jeweiligen Substrukturen und Bildschirmdokumenten erzeugt und diese in einer für das Juristenauge tauglichen Form repräsentiert. Sofern MS Internet Explorer 5.0 oder höher benutzt wird, kann auch mit einer einigermaßen brauchbaren Druckausgabe gerechnet werden. Besondere Erwartungen an die Darstellung im Web im Vergleich zu einer klassischen Papierausgabe stellen die Nutzer im Bereich der Verfügbarkeit von Rechtstexten zu alten Geltungsständen und allgemein im Bereich

der Verlinkung. Suchmasken zum Aufruf von zitierten Referenzstellen werden als Affront gewertet. Ein Mausklick auf ein Stichwort muss genügen, um Hintergrundinformation zu bekommen.

Kommen wir nun zur typographisch korrekten Ausgabe auf Papier mit Seitenformatierung. Mitarbeiterfortbildung im Bereich „electronic publishing“ und die Entwicklung elektronischer Produkte, welche die rentable Printauflage schmälern, stellen erhebliche Kostenfaktoren dar. Daher lohnt sich jede Erforschung von Einsparpotenzial sowohl für das produzierende Gewerbe (Verlagswesen) als auch für die haushaltsrechtlich gebundene öffentliche Hand. Solches Einsparpotenzial zeichnet sich ab. XML macht dies im Zusammenspiel mit Java möglich.

2. Voraussetzungen

- Die zu setzenden Daten müssen in wohlgeformtem XML vorhanden sein.¹
- Java Runtime Environment 1.3 muss verfügbar sein.²
- Die Klassenbibliothek „FOP“ von <http://xml.apache.org/fop> muss aktiviert sein.³

¹ Definitionen und Beispiele im Internet außer bei xml.apache.org vor allem bei xml.com und bei msdn.com. Eine XML-Datei sieht zB wie folgt aus:

```
<?xml version="1.0" encoding="UTF-8"?> <root> <entscheid> <head> Metzgerin-
nungsgericht Kleinbratlingen, Urteil vom 22.11.1998</head> <leitsatz> Es ist un-
schicklich, den Daumen auf die Waage zu legen, wenn es der Kunde be-
merkt.</leitsatz> <stichworte> <stichwort>Wurst</stichwort> </stichworte> <voll-
text>Fleisch Schweinebraten Service</volltext></entscheid>
<entscheid> <head>Ehrengericht der Adobe Corporation, Urteil vom
20.10.2001</head> <leitsatz>Es ist unschicklich, gratis Programme zu machen, die
dasselbe tun wie kostenpflichtige Software.</leitsatz> <stichworte> <stich-
wort>Acrobat Distiller</stichwort> <stichwort>PostScript</stichwort> </stichworte>
<volltext>Unicode Java Apache Foundation James Tauber</volltext> </entscheid>
</root>.
```

² ZB bei <http://java.sun.com/j2se/1.3/jre/> herunterzuladen oder auf Beipack-CD-ROMs zu aktuellen Computerzeitschriften erhältlich. Die gepackten Einzelmodule brauchen nicht entpackt zu werden. Es genügt ein PATH-Eintrag für das /bin-Verzeichnis der `java.exe` in der `Autoexec.bat`, um Java unter windows98 lauffähig zu machen. Auch die Entwicklerumgebung JDK ist für die Ausführung von Fop geeignet.

³ Diese Programmsammlung („Klassenbibliothek“) von *James Tauber* und Mannschaft kann unter apache.org/fop geladen werden und muss in der `Autoexec.bat` mit dem Eintrag „set CLASSPATH=C:\jars\Fop-0.20.1\build\fop.jar; C:\jars\Fop-0.20.1\lib\xerces-1.2.3.jar; C:\jars\Fop-0.20.1\lib\xalan-2.0.0.jar“ dem System einmalig bekannt gemacht werden.

- Ein XSL-Stylesheet, welches hinreichend Templates für die SML-Instanzen enthält, muss verfasst sein.⁴
- Das Stylesheet muss pro template die gewünschten output-Parameter in der xml:FO - Schreibweise enthalten.⁵
- Die Systemvariable JAVA_HOME muss auf JRE verweisen.⁶
- Die verwendeten Fonts müssen auf dem System verfügbar, lizenziert und in der Datei userconfig.xml bezeichnet sein.⁷

3. Methode

Auslösen des folgenden Batch-files:

```
fop.bat = java org.apache.fop.apps.Fop (mit folgenden
Parametern:)
-c conf/userconfig.xml
-xml Quelldatei.xml
-xsl Stylesheet.xsl
-pdf Ausgabedatei.pdf
```

4. Ergebnisse

Die zu setzenden Daten sind gesetzt und stehen als druckreifes PDF zur Verfügung.

⁴ Ein Stylesheet für die obige XML-Datei könnte zB so ein Skelett haben:

```
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0"
xmlns:fo="http://www.w3.org/1999/XSL/Format">
<xsl:template match =,"root">
<fo:root xmlns:fo="http://www.w3.org/1999/XSL/Format">
<xsl:apply-templates/>
</fo:root>
</xsl:template>
</xsl:stylesheet>
```

⁵ Zum Beispiel für <stichwort> und <volltext> aus der obigen XML-Datei:

```
<xsl:template match =,"stichworte/stichwort">
<fo:block font-size="9pt" font-family="sans-serif"
line-height="10pt" text-align="start" start-indent="1cm">
<xsl:value-of select=","/>
</fo:block>
</xsl:template>
<xsl:template match =,"entscheid/volltext">
<fo:block font-size="11pt" line-height="13pt"
text-align="start" background-color="white" color="black" >
<xsl:value-of select=","/>
</fo:block>
</xsl:template>
```

⁶ Das ist normalerweise nach der Java-Installation automatisch der Fall. Vergleiche <http://userpage.fu-berlin.de/~ram/java/installation.html>.

⁷ In der Datei C:\jars\Fop-0.20.1\confuserconfig.xml nach folgendem Muster:

```
<font metrics-file="arial.xml" kerning="yes" embed-file="arial.ttf">
<font-triplet name="Arial" style="normal" weight="normal">
<font-triplet name="ArialMT" style="normal" weight="normal">
</font>
```

4.1. Vorteile

- *kostenlos*, denn sowohl die Java- als auch die XSL-T-Technologie werden lizenzfrei von ihren Erfindern im Internet zur Verfügung gestellt und sogar in Anleitungsschriften erklärt.
- *open source*, denn es handelt sich einerseits um Dateien, die in ASCII nach der inhaltbestimmten XML-Strukturierung vorgehalten und von jedem Textverarbeitungsprogramm auf allen Betriebssystemen gelesen werden können, und andererseits bei FOP um Programme, die in dokumentierte Elemente und Module aufgeteilt sind und auf der plattformunabhängigen Java-Runtime ablauffähig sind.
- *zukunftsicher*, denn aufgrund der ersten beiden Parameter ist es gewährleistet, dass stets Programmierer zur Verfügung stehen, die systemimmanente Weiterentwicklungen leisten können; es gibt keine Zentrale, die in Konkurs fallen und den Source-Code mitnehmen kann.
- *browserfähig*, dh für den Benutzer stehen mehrere kostenlose Anzeigeprogramme (wie zB MSInternet Explorer) zur Verfügung, die den Inhalt der strukturierten Daten mit Bildschirmtypographie versehen und so erläuternd darstellen können.
- *satzkostenfrei*, denn FOP erledigt den systematisch korrekten Umbruch der XML-Datenbasis auf ein gegebenes Papierformat selbstständig gemäß einer ein Mal zu schreibenden Definition.

4.2. Nachteile

Ein Mitglied der sparwilligen Organisation muss XSL können. Es gibt keine Eingriffsmöglichkeit außerhalb des automatisch erzeugten Ergebnisses, Korrekturen müssen am Stylesheet oder an den XML-Daten ausgeführt werden, gelten dann für alle Fälle dieser Art und erfordern ein erneutes Durchlaufen des Rendering-Prozesses. Bei FOP sind außerdem noch nicht alle im xml:FO angelegten Formatanweisungen programmiert.

4.3. Prognose

Nach Ansicht des Autors wird Java-FOP (oder ein Derivat) im Laufe der nächsten Jahre die juristische Publikationswelt erobern und mit juristischen Publikationen befasste Satzbetriebe dazu zwingen, ihre Hauptaufgabe in Richtung Datenmanagement zu verlagern.