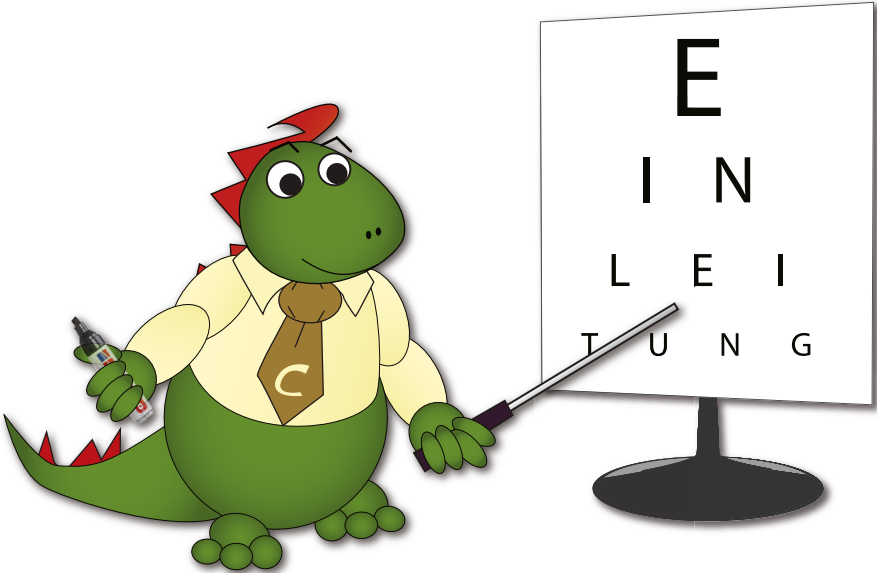


Vorwort



Vorwort

Herzlich willkommen zum zweiten Buch einer zweiteiligen Reihe zur Transformation von XML mit Hilfe von XSLT und XSL-FO, auch bekannt unter der Abkürzung XSL (eXtensible Stylesheet Language (for Transformations)). In diesem zweiten Band erfahren Sie alles Wesentliche, um mit Hilfe von XSL-FO 1.1 Algorithmen zu entwerfen, mit deren Hilfe XML-Dateien in PDF und andere Druckformate übertragen werden können. Darüber hinaus lernen Sie, wie Sie XPath und XSLT zusammen mit XSL-FO verwenden, um erweiterbare und konfigurierbare Umwandlungen zu erstellen.

Zielsetzung

XSL-FO (Formatting Objects) ist ein W3C-Standard, mit dessen Hilfe Sie Formatierungen für XML-Daten zur Druckausgabe erstellen können. Im Wesentlichen ist damit die Erzeugung von PDF-Dokumenten gemeint, aber in Abhängigkeit des Prozessors sind Sie auch in der Lage, andere Formate wie PostScript oder sogar HTML zu erzeugen. XSL-FO bietet dabei fast keine eigenen Möglichkeiten, Algorithmen zu entwerfen, sondern wird durch XSLT so erzeugt, dass Formatierungsanweisungen und Daten aus dem umgewandelten XML-Dokument in einer so genannten FO-Datei zusammen gebracht werden. Daher sind Kenntnisse von XSLT und XPath unerlässlich, um überhaupt geeignete Umwandlungen in das FO-Format zu schreiben. Dieses Buch konzentriert sich dabei ausschließlich auf die Darstellung der notwendigen FO-Elemente, um Seiten, Vorlagen und Block-Elemente sowie Zeichenformatierungen einzurichten. XSLT wird dabei grundsätzlich als bekannt vorausgesetzt.

Warum haben wir uns überhaupt für zwei Bücher entschieden? Es gibt bereits verschiedene Bücher zu XSLT und XSL-FO auf dem Markt, wie Ihnen sicherlich bereits aufgefallen ist. Auch im Internet finden sich verschiedene gut gemachte Anleitungen und Texte, wie mit XSLT und XSL-FO zu arbeiten ist. Natürlich möchten und werden wir uns inhaltlich von diesen Texten unterscheiden. Die Aufteilung in Bücher und Kapitel lässt sich so begründen, dass nicht jeder Leser beide Werke benötigt und er vielleicht eine Auswahl treffen möchte. Einige möchten möglicherweise auch nach und nach beide Bände kaufen und nicht von vorneherein ein erschreckend umfangreiches Buch in der Hand halten.

- Der erste Band stellt die Syntax von XSLT dar. Es umfasst die Standards von XSLT 1.0 mit dem Titel XSL Transformations (XSLT), Version 1.0, W3C Recommendation 16 November 1999 unter <http://www.w3.org/TR/xslt> und XPath 1.0 mit dem Titel XML Path Language (XPath), Version 1.0, W3C Recommendation 16 November 1999 unter <http://www.w3.org/TR/xpath> aus Version 1 und die Standards XSLT 2.0 mit dem Titel XSL Transformations (XSLT) Version 2.0, W3C Recommendation 23 January 2007 unter <http://www.w3.org/TR/xslt20/> und XPath 2.0 mit dem Titel XML Path Language (XPath) 2.0, W3C Recommendation 23 January 2007 unter <http://www.w3.org/TR/xpath20/>. Sie lernen in diesem Band daher alle Techniken kennen, welche zur Erzeugung von HTML, Text und XML und zur Abfrage von XML-Dokumenten mit XPath notwendig ist. Die Grundlagen werden im ersten Teil geschaffen. Der zweite Teil dieses Bandes stellt neben fortgeschrittenen Techniken eine Reihe von Einsatzbeispielen vor. Er kombiniert die einzelnen vorgestellten Standards mit allgemeinen Techniken, die nur die Transformation betreffen, mit Ideen, wie XSLT-Anwendungen aufgebaut werden können, und solchen Techniken wie die Verwendung einer Datenbank, um XML- und sogar XSLT-Daten abzuspeichern und dynamisch für Transformationsabläufe zusammenzusetzen.
- Der zweite Band stellt die Syntax von XSL-FO 1.1 bzw. XSL 1.1 mit dem Titel Extensible Stylesheet Language (XSL) Version 1.1, W3C Recommendation 05 December 2006 unter <http://www.w3.org/TR/xsl/> dar. Sie lernen in diesem Band, wie Druckerzeugnisse wie z. B. PDF-Dateien aus XML-Daten erzeugt werden. Dies erfordert einen eigenen Standard mit einer Vielzahl an unterschiedlichen Elementen, um Seitenbereiche, Seitenverläufe, Text- und Absatzformate sowie zusätzliche Dokumenteigenschaften anzugeben wie Inhaltsverzeichnisse.
- Weitere Bände zur XML-Verwendung in Software sind ebenfalls bei Comelio Medien erhältlich. Sie betreffen neben XSLT auch die Umwandlung und Erstellung von XML in Programmiersprachen sowie die Verwendung von XML in den Datenbanken MS SQL Server und Oracle.

Jedes Buch steht für sich. Sie bilden jedoch zusammen eine Einheit, die sich insbesondere aus der übergeordneten Zielsetzung speist, Transformationen möglichst raffiniert zu erstellen, um gute Algorithmen zu entwerfen.

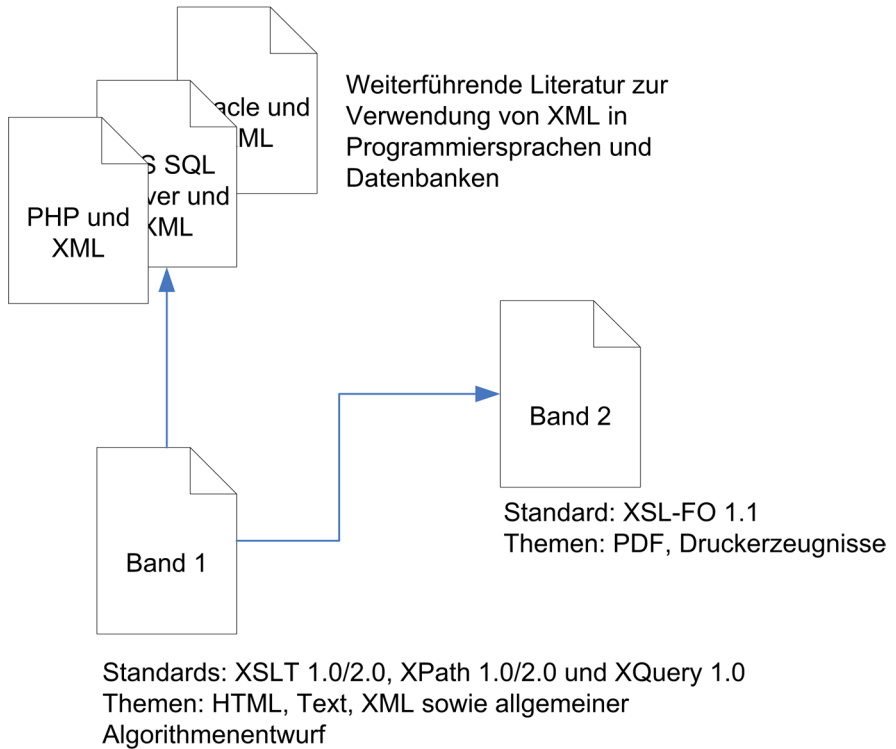


Abbildung v.1: Aufbau der Reihe

Leserkreis

Die zwei Bücher richten sich an Leser, die in jedem Fall bereits grundlegende Kenntnisse über XML und dessen Einsatzbereiche haben. Für diesen zweiten Band ist es empfehlenswert, sich vorab schon einmal mit XSLT zu beschäftigen – am besten natürlich mit unserem ersten Buch. Wie schon erwähnt, verhält es sich so, dass FO ein Vokabular für die Formatierung und nicht etwa eine Programmiersprache für die Umwandlung in Druckformate ist. Man erstellt mit Hilfe von XSLT aus einer XML-Datei eine FO-Datei, welche die Formatierungsangaben und die Daten enthält. Letztendlich könnte man also auch mit dem DOM oder mit sonstigen XML-Technologien einer beliebigen Programmiersprache eine solche FO-Datei erstellen und diese dann zu einem FO-Prozessor schicken, der aus diesem Dokument bspw. ein PDF erzeugt. Im Normalfall wird man aber XSLT verwenden, da beide Technologien optimal zusammen passen. Daher ist es unerlässlich, gute Kenntnisse von XSLT zu besitzen, da FO auf XSLT und XPath aufbaut.

Eine Präferenz für einen bestimmten Prozessor benötigen Sie nicht. Dieses Buch will auch keine Werbung für einen (kommerziellen) Prozessor machen, sondern im Wesentlichen die standardisierte Syntax von FO darstellen. Man wird dann bei Bedarf einen geeigneten FO-Prozessor verwenden und sich hier in der jeweils aktuellen Dokumentation erkundigen müssen, welche zusätzlichen proprietären XML-Elemente noch verfügbar stehen. Im Wesentlichen ist allerdings auch der Apache FOP-Prozessor als Open Source-Lösung ein Werkzeug, mit dem sich eine Vielzahl an Standardanforderungen umsetzen lassen. Das Buch hat daher die Zielsetzung, Sie mit der FO-Syntax vertraut zu machen und Ihnen dann selbstverständlich die Entscheidung zu überlassen, welche Software Sie letztendlich für die Ausführung einsetzen.

Autoren

Ein solches Werk schreibt man nicht alleine, sondern teilt sich die Arbeit so auf, wie persönliche Interessen und Wissen vorhanden sind. Die Autoren arbeiten in verschiedenen Positionen bei der Comelio GmbH (www.comelio.com) und haben bereits gemeinsam oder alleine verschiedene Bücher zu ihren Werkzeugen veröffentlicht.

- Marco Skulschus arbeitet als Projektleiter und Dozent und beschäftigt sich mit Berichts- und Expertensystemen auf Basis von MS SQL Server und Oracle. Er interessiert sich besonders für Ontologien und Data Mining-Techniken. Im XML-Bereich beschäftigt er sich mit der Abbildung von XML-Strukturen für Dateien und Datenbanken.

- Marcus Wiederstein arbeitet als Projektleiter für Projekte mit Microsoft-Produkten wie MS BizTalk Server, MS Sharepoint Portal Server oder MS Project Server sowie den Business Intelligence-Produkten. Im XML-Bereich ist er insbesondere an Import-/Exportschnittstellen interessiert.

Aufbau

Das Buch zeigt die verschiedenen Bereiche der Dokumenterstellung mit den XML-Elementen und –Attributen von XSL-FO. Dabei steht als letztes Ziel die Erstellung von großen Dokumenten im Vordergrund, wobei sozusagen alle einzelnen Hilfsmittel zusammen kombiniert werden müssen.

1. Grundkonzept: Das erste Kapitel erläutert, welche Prozesse bei der Umwandlung von XML nach FO und schließlich nach PDF geschehen. Man erfährt, wie die einzelnen Schritte bei einer Umwandlung unterschiedliche Zwischenergebnisse erzeugen und wie man schließlich das fertige Ergebnis erhält.
2. Vorlagen: Das zweite Kapitel stellt die verschiedenen Seiten- und Seitenverlaufsvorlagen mit ihren Eigenschaften wie druckbarer Bereich und (un)druckbare Randereiche sowie ihre Beziehungen untereinander dar.
3. Blöcke: Das dritte Kapitel erklärt, welche Blockstruktur für FO benutzt wird und welche allgemeinen Eigenschaften sich für alle Block-Elemente in FO ergeben. Dies betrifft die Bereiche Abstand, Auffüllung, Einzug, Rahmen und Hintergrund.
4. Tabellen: Das vierte Kapitel zeigt das in FO verwendete Tabellenmodell und zeigt neben den Elementen, die man für die Erstellung einer Tabelle benötigt, auch die Formatierungsmöglichkeiten über die Block-Eigenschaften.
5. Listen: Das fünfte Kapitel konzentriert sich darauf, das Listenmodell aus FO und die entsprechenden Formatierungsmöglichkeiten für Listen darzustellen.
6. Zeichenformatierung: Das sechste Kapitel verlässt die Block-Ebene und stellt die umfangreichen Optionen bei der Zeichen- und In-Zeilenformatierung vor: Schriftart, Farbe, Größe, Schnitt.

7. **Grafik:** Das siebte Kapitel erläutert, wie man Grafiken in ein Dokument einbettet, sie skaliert und ausrichtet oder wie man sie in den Hintergrund bringt und dort kacheln kann.
8. **Bucherstellung:** Das achte Kapitel schließlich zeigt verschiedene neue Techniken, mit denen große und profesisonelle Dokumente erzeugt werden können. Dabei werden bereits an anderer Stelle vorgestellte Vorgehensweisen noch einmal zusammengeführt, Werte mit Hilfe von XSLT in Variablen, Attributgruppen oder temporären Bäumen gespeichert und schließlich mit Inhaltsverzeichnis und internen Links so kombiniert, dass man eine gut erweiterbare und modulare Umwandlungsstrategie einrichtet.

Weitere Lern-Ressourcen

Online-Material vom Verlag:

Alle Übungsdateien in XML, XSLT, XSL-FO und entsprechende Ergebnisse finden Sie auf der Verlagsseite www.comelio-medien.com. Um die Ergebnisse von Transformationen in HTML zu testen, brauchen Sie nur einen aktuellen, XSLT-fähigen Internet-Browser, in dem Sie die XML-Dateien öffnen. Alternativ können Sie die sehr weit verbreitete Software von Altova XMLSpy® (Testversion kostenlos und erhältlich) von www.altova.com verwenden. Auch das MS Visual Studio erlaubt die Verwendung von XSLT. Alternativ können Sie Ihre Skripte auch in einem beliebigen Text-Editor erstellen und sie lediglich mit einem .NET/Java-XSLT-Prozessor oder der Programmiersprache Ihrer Wahl umwandeln. Diese Software erlaubt es Ihnen, die Umwandlung von XML in FO zu testen. Um allerdings auch noch die eigentliche PDF-Umwandlung durchzuführen, benötigen Sie einen entsprechenden Prozessor: bei der Darstellung der Grundlagen werden Prozessoren noch erwähnt. Hier liegt eine Reihe von verschiedenen Implementierungen vor, welche die Arbeit mit FO leider an den Einsatz von JavaScript oder SQL erinnern, weil hier unterschiedliche konkrete Umsetzungen auch unterschiedliche Erweiterungen bieten.

Weitere Bücher von Comelio Medien:

- XSLT, XPath und XQuery, Marco Skulschus, Marcus Wiederstein, Sarah Winterstone, ISBN 978-3-939701-50-7
- XML Schema, Marco Skulschus, Marcus Wiederstein, Sarah Winterstone, ISBN 978-3-939701-54-5

- PHP und XML, Marco Skulschus, ISBN 978-3-939701-00-2
- Oracle, PL/SQL und XML, Marco Skulschus, ISBN 978-3-939701-49-1

Online-Material vom W3C:

- XSL Transformations (XSLT), Version 1.0, W3C Recommendation 16 November 1999 unter <http://www.w3.org/TR/xslt>
- XML Path Language (XPath), Version 1.0, W3C Recommendation 16 November 1999 unter <http://www.w3.org/TR/xpath>
- XSLT 2.0 mit dem Titel XSL Transformations (XSLT) Version 2.0, W3C Recommendation 23 January 2007 unter <http://www.w3.org/TR/xslt20/>
- XML Path Language (XPath) 2.0, W3C Recommendation 23 January 2007 unter <http://www.w3.org/TR/xpath20/>
- Extensible Stylesheet Language (XSL) Version 1.1, W3C Recommendation 05 December 2006 unter <http://www.w3.org/TR/xsl/>
- Extensible Markup Language (XML) 1.0 (Third Edition), W3C Recommendation 04 February 2004 unter <http://www.w3c.org/TR/2004/REC-xml-20040204/>
- XML Schema Part 1: Structures, W3C Recommendation 02 May 2001 unter <http://www.w3c.org/TR/xmlschema-1/>
- XML Schema Part 2: Datatypes, W3C Recommendation 02 May 2001 unter <http://www.w3c.org/TR/xmlschema-2/>

Textgestaltung

Wie in allen Informatik-Büchern, finden Sie auch in diesem verschiedene typografische Hervorhebungen für *wichtige Begriffe* oder Quelltext sowie Internetadressen oder Dateinamen.

Quelltextblöcke haben besondere Formatierungen.

Kontakt zu Autoren und Verlag

Sie erreichen die Autoren an Ihrem Arbeitsplatz unter der E-Mail-Adresse info@comelio.com. Die Webseite des Verlags finden Sie unter der Adresse <http://www.comelio-medien.com>.

Auch in der realen Welt ist der Verlag zu erreichen. Die Hauptzentrale befindet sich in Berlin, Deutschland.

Comelio GmbH
Goethestr. 34
13086 Berlin