

Professionelle XML-Verarbeitung in Word



Manuel Montero Pineda studierte Geschichte, Psychologie und Erziehungswissenschaften in Heidelberg. Nach einem Aufbaustudium in Informatik an der FH Heidelberg arbeitete er als Berater und Entwickler für verschiedene Unternehmen im Bereich der technischen Dokumentation und der Druckvorstufe und gründete später die Data2type (www.data2type.de). Sein Tätigkeitsbereich umfasst alle relevanten XML-Technologien im Bereich des Publizierens. Bekannt wurde er als Autor diverser Artikel und des Buches "XSL-FO in der Praxis".



Jürgen Sieben ist Geschäftsführer der ConDeS, die sich seit 1999 mit der Entwicklung datenbankgestützter Applikationen und der Office-Integration für den Gesundheitsmarkt beschäftigt. Schwerpunkte der Arbeit sind Oracle-Datenbanken, XML und XSLT und das .NET-Framework sowie die Anbindung von kundenspezifischen Lösungen an bestehende Datenbank-Applikationen. Zudem ist er ein langjährig erfahrener Trainer für die Bereiche Oracle, XML/XSLT und .NET.

Manuel Montero Pineda · Jürgen Sieben

Professionelle XML- Verarbeitung in Word



dpunkt.verlag

Manuel Montero Pineda
Data2Type
montero@data2type.de

Jürgen Sieben
ConDeS GmbH & Co. KG
j.sieben@condes.de

Lektorat: René Schönfeldt
Copy-Editing: Ursula Zimpfer, Herrenberg
Satz: Data2Type, Mannheim
Herstellung: Birgit Bäuerlein
Umschlaggestaltung: Helmut Kraus, www.exclam.de
Druck und Bindung: Koninklijke Wöhrmann B.V., Zutphen, Niederlande

Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;
detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

ISBN-10 3-89864-415-4
ISBN-13 978-3-89864-415-0

1. Auflage 2007
Copyright © 2007 dpunkt.verlag GmbH
Ringstraße 19 B
69115 Heidelberg

Die vorliegende Publikation ist urheberrechtlich geschützt. Alle Rechte vorbehalten. Die Verwendung der Texte und Abbildungen, auch auszugsweise, ist ohne die schriftliche Zustimmung des Verlags urheberrechtswidrig und daher strafbar. Dies gilt insbesondere für die Vervielfältigung, Übersetzung oder die Verwendung in elektronischen Systemen. Es wird darauf hingewiesen, dass die im Buch verwendeten Soft- und Hardware-Bezeichnungen sowie Markennamen und Produktbezeichnungen der jeweiligen Firmen im Allgemeinen warenzeichen-, marken- oder patentrechtlichem Schutz unterliegen. Alle Angaben und Programme in diesem Buch wurden mit größter Sorgfalt kontrolliert. Weder Autoren noch Verlag können jedoch für Schäden haftbar gemacht werden, die in Zusammenhang mit der Verwendung dieses Buches stehen.

5 4 3 2 1 0

Einleitung

Word 2003 — für viele Benutzer ist dies sicher nur ein normales Update wie viele Updates vorher auch schon. Für andere ist diese Version jedoch Ausdruck eines Paradigmenwechsels bei Microsoft. Warum? Waren bislang die Office-Produkte Inbegriff von proprietärer Technik, macht Microsoft nun die interne Datenstruktur mit einer gut dokumentierten und auf Standards basierenden XML-Struktur jedermann zugänglich. Dadurch öffnen sich die Office-Dokumente der elektronischen Weiterverarbeitung und reduzieren die Abhängigkeit vom Originalprogramm. Genauso einfach ist es möglich, bestehenden XML-Inhalt mit Hilfe eines Stylesheets in eine Word-Datei zu überführen, in der der Inhalt anschließend angesehen, editiert und gespeichert werden kann. Mehr noch: Erstmals erlaubt es Word, dass Dateien eine durch das Programm überprüfbare Struktur erhalten. Dies werden vor allem Verlage und Unternehmen schätzen, deren technische Dokumentation, Buchprojekte oder Vertragstexte bislang von unterschiedlichen Editoren in relativer gestalterischer Freiheit in Word erstellt wurden.

Dieses Buch führt in die Unterstützung von XML in Office ein. Es zeigt auf, wie die neue Technologie zu nutzen ist und welche Szenarien sich daraus ergeben. Es erläutert das Schema zu WordML (Wordprocessing Markup Language), auf dem die XML-Dateien, die aus Word erzeugt werden, basieren, und zeigt Möglichkeiten des effizienten Umgangs mit diesen Dateien auf. Ebenso werden neue Technologien, die auf der XML-Unterstützung basieren oder eng mit ihr zusammenarbeiten, dargestellt, wie z.B. die SmartDocument-Solution oder die Visual Studio Tools for Office. Ziel dieses Buches ist es, dem ernsthaften Anwender und dem Programmierer eine Quelle an die Hand zu geben, um erfolgreich XML in Word für sich zu nutzen. Eine Referenz über WordML schließt das Buch ab.

An wen richtet sich dieses Buch?

Dieses Buch wurde für Entwickler, Entscheider und technisch fortgeschrittene Anwender geschrieben, die sich mit den neuen Möglichkeiten von XML in Word vertraut machen wollen.

Entscheider Entscheidern erläutert dieses Buch die Prinzipien, aber auch die Grenzen dieser Technologie und unterstützt so die strategische Planung für die Integration dieser Technik in eigene Projekte.

Entwickler Programmierer profitieren von der kommentierten Referenz zu WordML ebenso wie die kritische Besprechung der neuen Programmierparadigmen wie den SmartDocument-Solutions. Dies macht das Buch zu einem ständigen Begleiter bei der Entwicklung eigener Anwendungen, die XML integrieren.

Anwender Dem technisch versierten Anwender eröffnet sich durch die einfache Erstellung von Word-Dokumenten mit Unterstützung von XML, aber auch durch die Erweiterung der Makroprogrammierung um die SmartDocuments ein Weg, bestehende Lösungen zu erweitern und durch die Integration offener Standards effizienter zu machen.

Für den normalen Anwender ist die XML-Unterstützung im Regelfall weniger wichtig, da sie auf einer Ebene unterhalb des normalen Umgangs mit Word angesiedelt ist. Ein Randthema bleibt sie auch für Anwender und Entwickler, die bereits in der Vergangenheit ausschließlich mit Office-Produkten gearbeitet haben, obwohl auch hier z.B. durch die Integration von XML-basierten Web-Services oder die Möglichkeit, eigenen XML-Inhalt ohne die Formatierungsangaben aus Word zu exportieren, neue Anwendungsfelder eröffnet werden können.

Aufbau des Buches

Vgl. Kapitel 1, S. 3 Das Buch beginnt in Kapitel 1 mit einer Einführung in die neuen Fähigkeiten von XML in Word und zeigt die standardmäßige Unterstützung von XML in Word 2003. Anhand von Beispielen lernt der Leser die Funktionalitäten Schritt für Schritt kennen.

Vgl. Kapitel 2, S. 27 Das Einführungskapitel 2 zu WordML, dem neuen Dokumentenformat von Word 2003, erläutert ausführlich diese nicht immer leicht zugängliche Auszeichnungssprache WordML. Der Leser lernt anhand von Beispielen alle wichtigen Zusammenhänge von WordML kennen. Dieses Kapitel bildet die Grundlage für die späteren Kapitel, in denen praktische Anwendungsszenarien entwickelt werden.

Vgl. Kapitel 3, S. 77 Eine auf die Anforderungen von WordML zugeschnittene Einführung in XSLT (eXtensible Stylesheet Language Transformations), dem W3C-Standard für die Transformation von XML-Daten, beschreibt die Sprache mit einfachen Beispielen. Die in den Kapiteln 3 und 4 erlernten Techniken werden im darauf folgenden Anwendungsteil dieses Buches immer wieder verwendet.

Der Anwendungsteil des Buches geht auf die in der Praxis üblichen Workflows im Zusammenhang von Word und XML ein. Es wird detailliert beschrieben, wie aus eigenen XML-Strukturen mittels XSLT

WordML-Dokumente automatisch generiert werden können und wie WordML-Dokumente ebenfalls mit XSLT in eigene XML-Dateien umgewandelt werden können.

Grundlage für die hier gezeigten Beispiele bildet die weit verbreitete DocBook-Struktur, die in Kapitel 6 in einer vereinfachten Version dargestellt wird.

Vgl. Kapitel 6, S. 125

Den Themen SmartDocuments (Kapitel 9 und 10) und Visual Studio Tools for Office (Kapitel 11) sind eigene Kapitel gewidmet.

SmartDocuments sind eine neue Vision, die erst mit den Mitteln von XML in Word möglich wurden. SmartDocuments versprechen eine kontextbezogene Palette von Hilfsmitteln und Steuerelementen, mit deren Hilfe bislang nicht mögliche Benutzerführungskonzepte in Word umgesetzt werden können. Anhand eines Beispielprojekts wird die Arbeitsweise erläutert und Möglichkeiten und Grenzen der Technik aufgezeigt.

Vgl. Kapitel 9, S. 175

Die Zukunft der Anwendungsentwicklung mit Office stellen die Visual Studio Tools for Office (VSTO) dar, die das seit Einführung von Word bekannte, auf VBA (Visual Basic for Application) basierende Programmiermodell durch Managed Code des .NET-Frameworks ersetzt und bislang nicht gekannte Professionalität in der Programmierung von Word ermöglicht.

Der letzte Teil des Buches bietet eine Referenz zu WordML, XSLT, XPath und DocBook.

Website zum Buch

Alle im Buch vorkommenden Beispiele und Codes können von der Homepage des Buches unter <http://www.data2type.de/xml/word2003.html> heruntergeladen werden.

Danksagung

An dieser Stelle möchten wir allen danken, die uns bei der Erstellung dieses Buches unterstützt haben. Mein erster und ganz besonderer Dank geht an Steffen Herkert, der uns u.a. bei der nicht immer spannenden Arbeit an der Referenz unterstützt hat.

Ein großes Dankeschön geht ebenfalls an unsere Gutachter, die uns sehr wertvolle Hinweise gegeben haben, von denen viele in das Buch eingeflossen sind. Namentlich sind hier zu nennen Dr. Victor Wang (Verlagsgruppe Hüthig Jehle Rehm), Marko Hedler (pagina GmbH) und Ursula Welsch (Ursula Welsch Neue Medien).