



LIFE THROUGH MARS: XML BRINGT LEBEN INS PDF

Adobes Mars-Projekt vereint die Vorteile von XML und PDF, ohne die Nachteile zu übernehmen. So werden ganz neue, hochflexible Publishing-Prozesse möglich.

PDF und XML: Unterschiede

Das PDF (Portable Document Format) ermöglicht die plattformunabhängige Darstellung und Verarbeitung von Dokumenten. Vorteil: Die Inhalte werden überall exakt so dargestellt, wie sie in der Quell-Applikation formatiert wurden. Mitunter ein Nachteil: Grafiken, Bilder, die Typografie und Farben sind statisch integriert. Das bedeutet, in einem PDF-Dokument sind der Content und die Darstellung stark aneinander gekoppelt.

Die XML (eXtensible Markup Language) funktioniert genau entgegengesetzt. Sie trennt den Inhalt von der Visualisierung. Dabei zeichnet XML den Content mit sogenannten Tags aus. Wie der Inhalt schließlich dargestellt wird, hängt einzig und allein von der verarbeitenden Applikation und der Zielpattform ab. Vorteil: Ein XML-Dokument kann sowohl in einem Web-Browser als auch in einem Office-Programm und in einer professio-

nellen Layout-Applikation wie InDesign funktionieren. Seine Darstellung wird aber – im Gegensatz zu PDF – überall unterschiedlich sein.

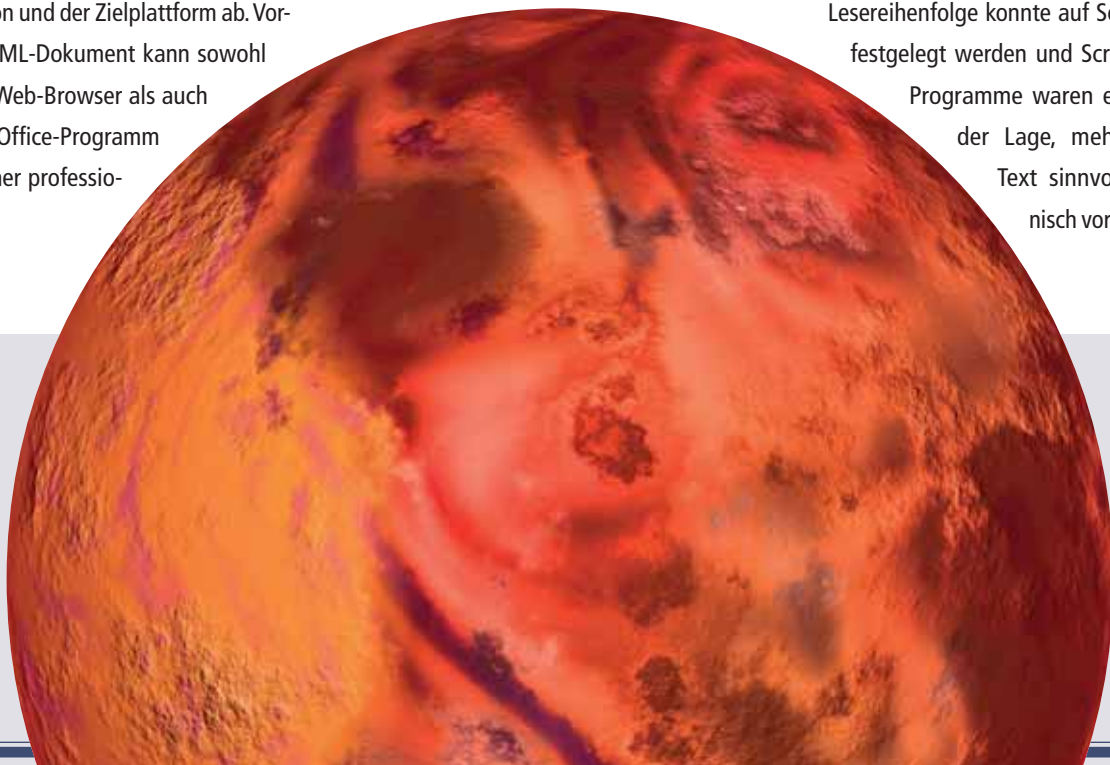
Die ersten XML-Strukturen in PDF

In der Entwicklungsgeschichte des PDF standen mit der Acrobat-Version 4 erstmals Formularwerkzeuge zur Verfügung. Damals wurden bereits die ersten XML-Strukturen eingeführt – in Form von fdf (Forms Data Format). Auf diese Weise konnte man den Inhalt, der in ein PDF-Formular eingegeben wurde, an einen Web-Server übertragen. Acrobat 5 erweiterte das fdf schließlich zu xfdf und machte es damit zum reinen XML. Die Acrobat-Version 5 stellte auch einen Fortschritt in der „kollaborativen Zusammenarbeit“ dar:

Abstimmungsprozesse konnten nun mit Kommentaren in PDF-Dokumenten abgewickelt werden – per E-Mail, WebDav und Browser. Die PDF-Kommentare wurden in XML codiert.

Auch die interne Repräsentation von Metadaten wurde in dieser Phase auf Standard-XML umgestellt. Die Bedeutung dieser XML-basierenden Metadaten wurde später in Acrobat 6 durch die Einführung von XMP massiv vergrößert. Dank XMP war es nun möglich, Urheberrechte, Kameradaten und viele Detailinformationen an Dokumente zu hängen.

Mit dem Zusatzmodul „Accessibility“ konnten bereits in Acrobat 5 Inhalte von PDF-Dokumenten mit Tags versehen werden, Bilder mit alternativen Texten. Die Lesereihenfolge konnte auf Seitenebene festgelegt werden und Screenreader-Programme waren erstmals in der Lage, mehrspaltigen Text sinnvoll elektronisch vorzulesen. U



Durch „tagged PDF“ war es endlich möglich, PDF-Dokumente auch auf mobilen Endgeräten korrekt abzubilden. Hierfür zeichnete die Umfließen-Eigenschaft verantwortlich: Sie ließ jetzt für PDF-Text, der ja rein zeilenorientiert und starr codiert ist, in anderen Anwendungen Umbrüche zu. Die eigentliche Stärke von „tagged PDF“ lag und liegt aber in den Acrobat-Exportfiltern, die PDF-Dokumente in andere Formate überführen. Nur „tagged PDF“ wird bei komplexen PDF-Inhalten sinnvoll nach RTF, DOC, XML und HTML überführt.

In der Acrobat 8 Professional-Version kann das Tagging von PDF-Content nachträglich vorgenommen werden.

XML und PDF: die aktuelle

Entwicklung

Von Acrobat 4 bis Acrobat 8 hat eine stetige Erweiterung der PDF-Sprache um XML-Elemente stattgefunden. Für viele Acrobat-User war das nicht unbedingt erkennbar. Warum auch: PDF ist seiner eigentlichen Bestimmung, plattformübergreifend Dokumente auszutauschen und für den Druck auszugeben, treu geblieben. Seit PDF1.7 (Acrobat 8) bestehen intelligente Dokumente aus XML-Strukturen im PDF-Container. Oder anders formuliert: PDF wird immer mehr zu einem XML-ba-

sierenden Format. Dieser Trend wird neben der Integration von PDF in Adobe LifeCycle auch im Mars-Projekt sichtbar.

Das Mars-Projekt

Vergleichbar mit den LiveCycle-Formularen setzt Adobes Mars-Projekt auf die hierarchische Abbildung von PDF-Elementen in einer reinen XML-Struktur. Die Vorteile von PDF für die sichere Darstellung und Reproduktion von Fonts, Farben und Bildern bleiben dabei voll und ganz erhalten. Die Transformationsmöglichkeiten von XML kommen hinzu und schaffen ganz neue Möglichkeiten, den Seiteninhalt von Dokumenten komplett dynamisch zu halten, ihn zu verändern oder gar anwendungsspezifisch neu zu erzeugen.

Eine Mars-Datei ist ein Zip-komprimierter Container, der neben den PDF-typischen Elementen auch skalierbare Vektorgrafiken (SVG) und XPath-Elemente enthalten kann. SVG kommt bei der Präsentation des Seiteninhalts zum Einsatz. Jede „PDF-Seite“ eines Mars-Dokuments wird durch eine skalierbare Vektordatei (SVG) dargestellt. XPath wird als Anfragesprache für XML-Transformationen benutzt. Mars könnte somit ganz neue Publishing-Prozesse ermöglichen, die auf Basis der

heutigen Technologien noch nicht denkbar sind. Es bleibt also spannend, wann Adobe die erste Alpha-Version, die dann gewiss unter anderem Namen erscheint, auf den Markt bringen wird.

Fazit

Mit Adobe Acrobat 8 Professional und „tagged PDF“ hat XML endgültig Einzug in den PDF-Standard gehalten. Ohne die „Accessibility“-Funktionen in Acrobat könnten PDF-Dokumente nicht sinnvoll in andere strukturierte Formate wie RTF, DOC, XML oder HTML konvertiert werden. Wer solche Umwandlungen benötigt und bisher enttäuscht worden ist, der sollte PDFs unbedingt einmal taggen, bevor er ans Exportieren geht.

Für künftige webbasierte Publishing-Lösungen stellt eine XML-Repräsentation von PDF die Basis für neue Systeme dar. Betrachtet man das Mars-Projekt als Teil anderer Adobe-Technologien, ist es nicht unwahrscheinlich, dass es in einer Kombination mit Flex, Flash und Air die Publishing-Welt revolutionieren könnte. Die Grenzen des Machbaren werden allein durch die Fantasie der Entwickler definiert. □ **Bernd Kress**



Bernd Kress

TimeToMedia
bernd.kress@t2media.de
www.t2media.de