

deutlich von den üblichen Strukturen (wie z. B. bei DocBook) unterscheidet, sodass eine direkte Nutzung dieser Struktur nicht empfehlenswert ist. Allerdings werden beim Speichern in WordML die Namen der verwendeten Absatzformate als Attribute mitgegeben (s. Abb. 1), sodass diese Formatvorlagen im WordML-Datenstrom gut zu identifizieren sind.

Zwar ist das WordML-Format auf den ersten Blick unübersichtlich und flach, aber man kann auf Grund der Aufwärtskompatibilität der verschiedenen Word-Versionen jedes ältere Word-Dokument ganz rasch und ohne Aufwand in dieses Format umsetzen. Da WordML für Skripte und Routinen gut greifbar ist, wird es zum Schnittstellenformat für die Weiterverarbeitung von älteren Word-Dokumenten – und damit zum Retter von größeren Datenbeständen, die auf eine neue Plattform gebracht werden sollen.

Benutzerdefiniertes XML

Jedes XML-Dokument lässt sich in Word öffnen – und zwar nicht nur als reine Textdatei wie in früheren Word-Versionen, sondern als XML-Datei mit grafisch dargestellten Tags. Es wird also automatisch zwischen textlichem Inhalt und Mark-up unterschieden. Damit kann MS Word 2003 jederzeit als XML-Editor eingesetzt werden.

Soll die XML-Struktur in Word 2003 bearbeitet werden, z. B. durch das Einfügen neuer Elemente, dann wird ein XML-Schema (W3C-Schema) benötigt, das die Elemente und Attribute bereitstellt und während der Eingabe die Validierung übernimmt. Zwar wird in der Technischen Dokumentation in der Regel mit Doku-

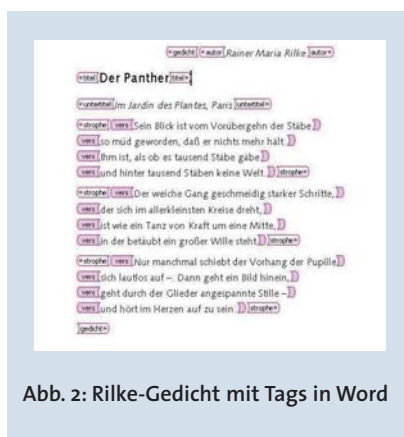


Abb. 2: Rilke-Gedicht mit Tags in Word

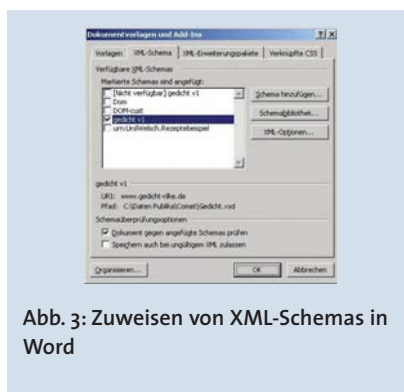


Abb. 3: Zuweisen von XML-Schemas in Word

menttyp-Definitionen (DTDs) gearbeitet, aber für DocBook und DITA gibt es vorbereitete XML-Schemas zum Download. Im Übrigen gibt es diverse Tools, die DTDs in XML-Schemas umsetzen und umgekehrt, Word selbst tut dies allerdings nicht. Die verschiedenen Schemas werden in einer Schema-Bibliothek Office-weit verwaltet. Dem Dokument wird das XML-Schema über den Befehl Extras/Vorlagen und Add-Ins bekannt gemacht (s. Abb. 3). Parallel zum fest implementierten WordML kann also ein XML-Dokument in Word 2003 z. B. auch der DocBook-Struktur entsprechen. Das ist die eigentliche Leistung, die Microsoft vollbracht hat

und deren Bedeutung bislang – wohl auch von Microsoft selbst – noch bei weitem unterschätzt wird.

Ausblick

So weit, so gut – aber wie geht es weiter? MS Office 12 wird Anfang nächsten Jahres erscheinen, die ersten Beta-Versionen sollen bereits diesen Herbst zur Verfügung stehen. Bei Microsoft wird wohl noch heftig daran gearbeitet, die internen XML-Strukturen auch als Dateiformat (im Zip-Container) zu nutzen. Darüber hinaus sollen die elend langen Style- und Property-Definitionen, die bislang noch das Auffinden der eigentlichen Information in WordML-Dateien erschweren, ausgelagert werden und alle OLE-Elemente, die noch als binäre Objekte den XML-Datenstrom bevölkern, als externe Dateien XML-üblich eingebunden werden. Als besonderes Schmankerl ist angekündigt, dass ältere Office-Versionen nachträglich mit Dateikonvertern aufgerüstet werden können, damit auch diese in den Genuss der neuen XML-Formate kommen.

Die Integration von XML in MS Word bzw. MS Office wird also mit großer Energie vorangetrieben. Es ist und bleibt spannend. Nichtsdestotrotz ist die bereits verfügbare XML-Funktionalität von MS Word 2003 in den gängigen Arbeitsumgebungen sinnvoll und kostensparend einsetzbar, auch wenn manch kantige Eigenschaft noch geglättet werden muss.

WILLI BREITWIESER, ENTWICKLER FÜR MS-OFFICE-ANWENDUNGEN IM XML-UMFELD, WILLI@BREITWIESER.NET

URSULA WELSCH, FACHBERATERIN UND TRAINERIN FÜR DIE ANWENDUNG VON XML, URSULA.WELSCH@WELSCHMEDIEN.DE