In eigener Sache

Die elektronische Publikation des CMS-Journals

Natalia Ivanova
natalia.ivanova@cms.hu-berlin.de

Christiane Schöbel
schoebel@cms.hu-berlin.de

Matthias Schulz
matthias.schulz.1@cms.hu-berlin.de

In vorangegangenen Ausgaben unserer Zeitschrift wurde oft das Thema des SGML/XML-basierten Publizierens in verschiedenen Facetten beleuchtet. Seit der Nr. 23 wurde die Erstellung dieser Zeitschrift selbst auf eine neue und XML-basierte Technologie umgestellt. Im folgenden Bericht möchten die Autoren den neuen Publikationsprozess vorstellen.

Projektziele

Die Umstellung einer seit 12 Jahren herausgegebenen Zeitschrift kann nicht von heute auf morgen passieren und muss gründlich geplant werden. Wir haben uns dabei folgende Ziele gestellt:

- Verbesserung der Weiterverarbeitung von Worddokumenten in QuarkXPress
- Verringung der manuellen Formatierung in QuarkXPress durch automatische Übernahme von Formatvorlagen aus Worddokumenten in QuarkXPress-Stilvorlagen. Die Arbeit des Setzens sollte nicht durch einen Medienbruch zwischen dem Text- und dem Satzprogramm behindert werden.
- Mit der Ausgabe 24 sollte ein flexibler Layoutwechsel ermöglicht werden.
- Die Artikel sollten nicht nur als PDF-Dateien im Internet veröffentlicht werden, sondern auch als HTML-Versionen.
- Die Recherche sollte nicht mehr beschränkt sein auf Autor und Titel, sondern bei einer Volltextsuche sollte qualifiziert nach einzelnen Bestandteilen des Textes gesucht werden.
- Die Langzeitverfügbarkeit der Artikel sollte gesichert sein.

War früher alles besser?

Seit der Erfindung des Buchdrucks ist es üblich, dass Autoren ihre Artikel einem Setzer übergeben und dieser gestaltet nach typographischen Regeln die fertige Publikation. Trotz der Einführung von Computern haben einige Arbeitsschritte überdauert, die auch Gutenberg vertraut waren. So setzte er nach einem Manuskript jeden Buchstaben einzeln. Heute haben sich diese Arbeiten nur wenig verändert und bis zum Heft 22 wurde ähnlich aufwändig am Computer gesetzt. Die Autoren reichten die Artikel als Microsoft Word-Dateien ein. Nach dem Import in das Satzprogramm QuarkXPress und der Zuweisung der Stilvorlagen waren alle Textauszeichnungen (fett, kursiv usw.) nicht mehr vorhanden. Diese mussten anhand des ausgedruckten Manuskriptes alle

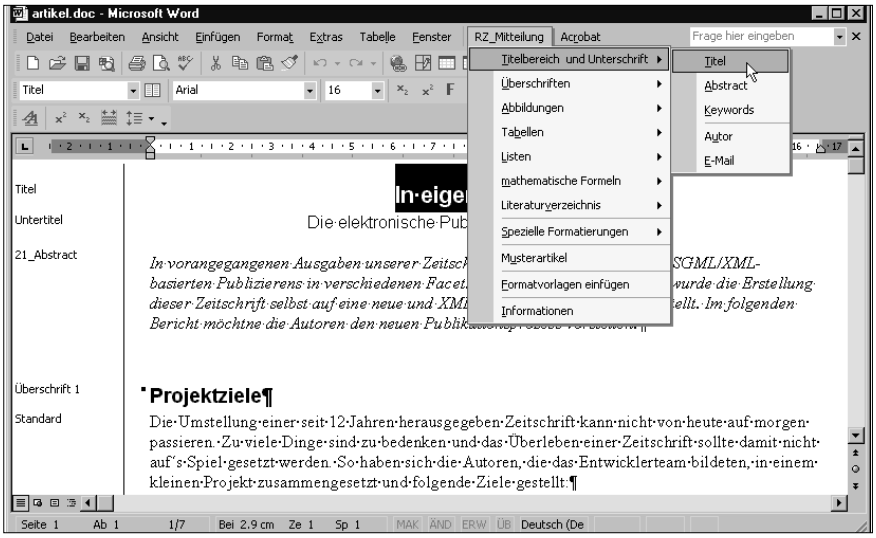

Abb. 1: Word-Ansicht mit der Formatvorlagenspalte (links) und dem aufgeklappten Menü RZ_Mitteilungen. Zeichenformate, wenn vorhanden, werden im Formatvorlagenfenster (in der Symbolleiste Format-Artikel) angezeigt.

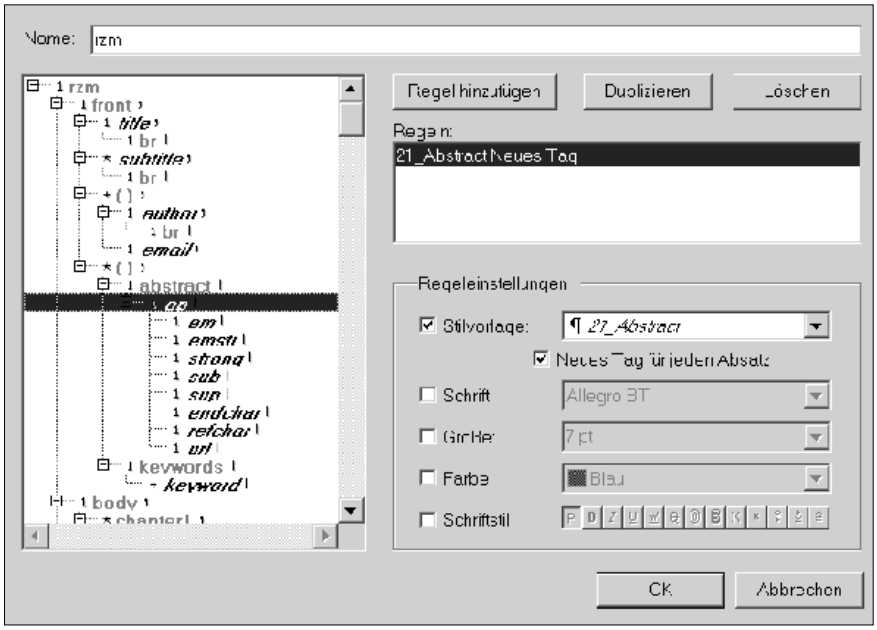

Abb. 2: In diesem Dialogfenster werden die Beziehungen zwischen den Stilvorlagen und den Tagelementen der DTD festgelegt.

manuell nachvollzogen werden. Es braucht nicht erwähnt zu werden, dass diese Prozedur nicht nur einen erheblichen Aufwand bedeutete, sondern auch eine große Fehlerquelle in sich barg.

Diese Arbeit wiederholte sich für jeden der ca. 15-20 Beiträge eines Heftes. Das Problem bei dieser Bearbeitung ist, dass der Text erst während des Setzens mit Stilvorlagen versehen wird. Formatierungen und die Struktur des Beitrages werden vom Satzprogramm nicht automatisch erkannt. Kurz gesagt, Formatierungen von Word fließen nicht in QuarkXPress-Stilvorlagen ein. Um dies überbrücken zu können, versehen die Autoren die Word-Dateien nun mit Formatierungsinformationen, sodass ein automatisches Setzen möglich wird.

Gut strukturiert ist halb gewonnen

Um eine einfache Lösung zu finden, wurde davon ausgegangen, dass die Autoren ihre Artikel mit Microsoft Word schreiben. Ein Gerüst mit Formatvorlagen sollte helfen, den Artikel mit Formatierungsinformationen zu versehen. Gestalterische Vorgaben treten damit in den Hintergrund, da es bei diesem Verfahren allein auf die Struktur des Textes ankommt.

Diese konzeptionellen Überlegungen mündeten in die Dokumentvorlage *rz-mit.dot*[1]. In dieser Dokumentvorlage befinden sich nicht nur die entsprechenden Formatvorlagen, sondern beim Schreiben wird dem Autor ein extra Menü *RZ_Mitteilungen* zur Verfügung gestellt. Somit wurde es ihm so einfach wie möglich gemacht, da alle notwendigen Formatvorlagen in diesem Menü vereint sind. Der Umgang mit Dokumentvorlagen ist jedoch sehr ungewohnt und das Wissen, welche Formatvorlage für welchen Absatz verwendet werden, soll kann sich nicht immer aus dem Kontext des Namens der Vorlage erschließen. Deshalb wurde von den Entwicklern ein multimediales Manual für den Umgang mit der Dokumentvorlage erstellt. [1]

Was hat sich mit der Einführung und Verwendung der Dokumentvorlage verändert? Der neue Workflow erfordert nun, dass die Autoren die Dokumentvorlage *RZ-Mit* für das Schreiben ihrer Artikel anwenden.

Mit Hilfe des Menüs *RZ_Mitteilung* sollen sie ihre Texte sorgfältig mit Formatvorlagen versehen. In der Realität sahen die Ergebnisse für die Autoren des Heftes Nr. 23 sehr heterogen aus. Der Import dieser Datei vereinfachte sich, da jetzt die Formatvorlagen von den Stilvorlagen übernommen werden konnten. Bei guter Vorarbeit der Autoren musste nur noch das Feinlayout ausgeführt werden. Das Setzen beschleunigte und automatisierte sich.

In der Bearbeitungsebene Word stellten sich jedoch technische Schwierigkeiten heraus, die das Entwicklerteam nicht voraussehen konnte. Diese waren unter anderem »neue eigene Formatvorlagen« die in Word erzeugt werden, dass jeder Autor auch neue Formatvorlagen erzeugen kann und Texte, die aus anderen Dokumenten mit fremden Formatvorlagen übernommen wurden. Einige Probleme konnten durch die korrekte Anwendung der Dokumentvorlage eingegrenzt werden, andere können zum jetzigen Zeitpunkt leider nocht nicht abgestellt werden. Auf der Ebene der Bearbeitung in QuarkXPress mussten zusätzliche und differenzierte Stilvorlagen erstellt werden. Dies war für den Export nach XML notwendig.

Der weitere Weg – entlang der Quark-Avenue zum XML

Warum XML

XML bietet heute die einzige Möglichkeit, die gesteckten Projektziele zu erreichen. Durch sein selbstbeschreibendes Textformat ist es möglich gezielt nach Informationen zu suchen, Transformationen nach Formaten wie HTML und PDF mühelos zu verwirklichen und für Langzeitverfügbarkeit dieses Formats zu sichern. Voraussetzung für XML ist, dass es vom Satzprogramm unterstützt werden muß.

Quark-Avenue

Seit der Version 4.0 ist es in QuarkXPress mit dem Zusatzprogramm *avenue.quark* möglich, XML-Instanzen entsprechend einer Dokumenttypdefinition (DTD) zu erstellen.

Dieses Tool, das in der aktuellen Version 5.0 von QuarkXPress integraler Bestandteil ist, extrahiert strukturierten Inhalt aus Dokumenten aufgrund der verwendeten Stilvorlagen.

Das Grundprinzip bei der Arbeit mit avenue.quark ist sehr einfach: ein Rahmen im Dokument wird ausgewählt und mit der Maus auf das entsprechende Element in der Liste *XML-Baum* auf der Palette *XML-Arbeitsbereich* geschoben. Falls der gesamte Fließtext sich in einem Textrahmen befindet, ist danach der größte Teil des Artikels getaggt, und es bleiben nur einige Sonderfälle wie Fußnoten u. ä.

1 Die alten Bezeichnungen in Dateien (*rz-mit.dot, rzm.dtd*) sind zz. beibehalten.
2 Bei der Entwicklung der Dokumentvorlage, sowie der DTD, konnte auf Erfahrungen der Projekte *Digitale Dissertation* (DiDi), *ProPrint* und *XML-Portal* der Arbeitsgruppe Elektronisches Publizieren zurückgegriffen werden. Hier wurden Konvertierungstechniken erarbeitet, die es ermöglichen, Word-Dateien direkt in SGML/XML umzuwandeln.
3 Countinuos Acquisition and Life-Cycle Support

übrig. Textteile, die jeweils in einem eigenen Rahmen platziert sind müssen in gleicher Art und Weise getaggt werden. Bilder, mathematische Formeln, Beispiele und Bildunterschriften sind ebenso gesondert zu behandeln.

Damit dies so möglich ist muß zuvor eine DTD modelliert werden. Der Aufbau der *RZM-DTD*[2] ist an die *DiML-DTD* (DissertationMarkupLanguage DTD für Dissertationen und Hochschulschriften) angelehnt. Spezielle Anforderungen und Strukturen wurden in einer empirischen Dokumentananlyse ermittelt und in der DTD definiert. Aber nicht alle modellierten Strukturen können in einem XMT-Formular in QuarkXPress umgesetzt werden. Somit wurde eine Zwischen-DTD festgelgt. Die mit dieser DTD erzeugten XML-Instanzen werden mittels eines XSLT-Stylesheets in das Format der RZM-DTD transformiert. Das erwähnte XMT-Formular wird in QuarkXPress erstellt. Es stellt die Verbindung zwischen den Stilvorlagen und den Elementen der DTD dar. (s. Abbildung 2).

„alles Quark?" – Probleme und Tücken

Trotz der leichten Handhabung des Programms gab es Schwierigkeiten, die zu umschiffen waren. So kann *avenue.quark* nicht zwischen Absätzen, die sich im Abstrakt oder im Kapitel befinden, unterscheiden, sondern übernimmt die einmal definierte Tagregel automatisch für alle gleichnamigen Elemente unabhängig von deren hierarchischer Struktur. Die Lösung war die Zwischen-DTD, die zusätzliche Tags für alle unterschiedlichen Absätze enthält.

Ein bis jetzt ungelöstes Problem stellt das Taggen von Tabellen dar. QuarkXPress unterstützt nicht das CALS[3]-Tabellenmodell wie FrameMaker 7.0 oder das HTML-Tabellenmodell. Somit müssen Tabellen von einem externen Konvertierungsprogramm erstellt werden.

Von XML zurück nach Quark

Trotz der dokumentierten und in der Fachlitertur beschriebenen Möglichkeit von QuarkXPress, auch XML-Dateien zu verarbeiten genügte uns dies nicht. Da das Verarbeiten von XML-Daten auf sogenannten Paltzhaltern bei QuarkXPress basiert, ist es nur möglich, feste Dokumentstrukturen in einem Dokument zu platzieren (z.B. wechselnde Adressen).

Was sollte realisiert werden? Ein Heft beinhaltet neben den Artikeln auch feststehende und dynamische Rubriken wie Impressum, Ansprechpartner und monatsaktuelle Weiterbildungsangebote. Diese Rubriken werden auch in anderen Kontexten wie den Webseiten der Weiterbildungsangebote der Universität, der Ansprechpartner des Computer- und Medienservice und der Impressumangabe der HTML-Präsentation benutzt. Um mit diesen Daten flexibel zwischen diesen Anwendungen mühelos wechseln zu können war XML die erste Wahl. So werden die Daten gleich im XML-Format gespeichert und editiert und für QuarkXPress und das Internet passend transformiert.

Da es sich bei diesen Daten nicht um Dokumente mit festen Strukturen handelte und deren Komplexität schon über der einer Adressenangabe steht war das Platzhalterkonzept nicht anwendbar. Hier

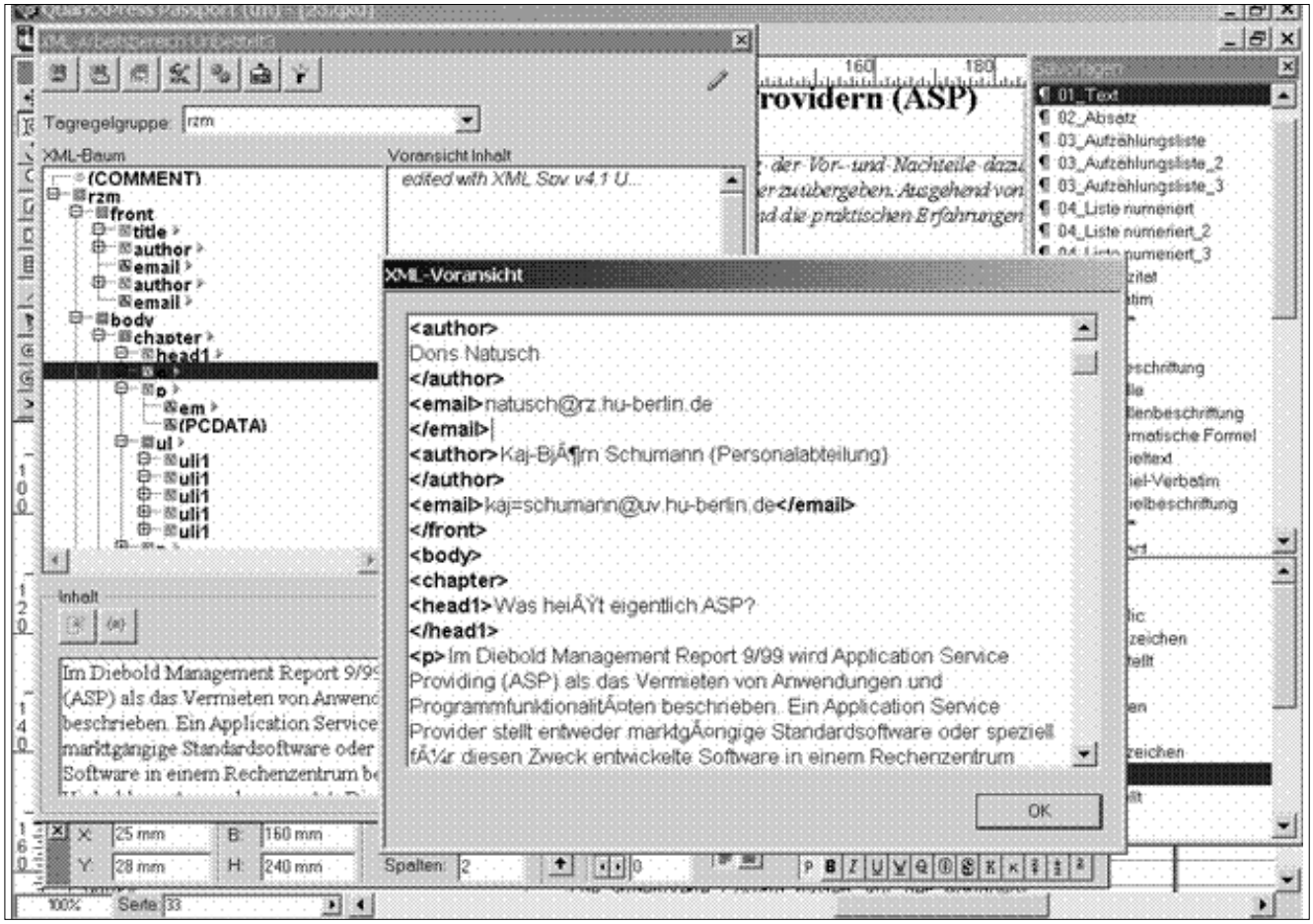

Abb. 3: QuarkXPress-Oberfläche mit Stilvorlagen (rechts), Palette XML-Arbeitsbereich (links) und XML-Voransicht (im Vordergrund).

erinnerten wir uns an die Möglichkeiten die XpressMarken boten und programmierten einen XSLT-Stylesheet, der die einzelnen XML-Dateien in das XTG-Format transformiert. Teile des Dokuments wie z. B. das Weiterbildungsangebot konnten so einfach und automatisiert layoutet in das laufende Heft eingesetzt werden.

„Ab ins Netz!"

Das Erzeugen von XML-Instanzen ist im gesamten Arbeitsprozeß ein nicht zu unterschätzender Teil. Doch hat man die XML-Dateien von allen Artikeln erst einmal vorliegen, kann man die Stärken von XML voll auspielen. Es ist unter anderem ein Leichtes, daraus HTML zu erzeugen, um die komplette Ausgabe im Internet zu präsentieren [2]. Die Transformation nach HTML geschieht mit einem XSLT-Stylesheet; es ist nur am Rande anzumerken, dass dessen Programmierung deutlich weniger Mühe kostete als die Einarbeitung in *avenue.quark*.

Ein Ausblick : Mit der Intergration des Dokumentenserver in ein Content-MangementSystem wird aktuell an einer dynamischen Transformation der XML-Daten gearbeitet.

Und was passierte bei Heft 24?

Mit einem neuen Namen und dem Umzug in ein neues Haus änderte sich auch die Gestaltung unserer seit 1991 herausgegebenen Veröffentlichungsreihe, ehemals *RZ-Mitteilungen* – nun *cms-journal*. War da alle Arbeit umsonst, die mit Heft 23 eingeführt wurde?

Das Paradigma von XML lautet Inhalt, Struktur und Layout von einander zu trennen. Das Spektrum und die Struktur bleiben auch beim Heft 24 und den folgenden bestehen. Ein neues Layout-Kleid wurde entworfen und geschneidert. Mit der Änderung der Stilvorlagen in QuarkXPress war die technische Umsetzung in wenigen Stunden vollzogen. Und aus einem 12-jährigen Kind wurde ein modern gestaltetes Computer- und Medienservice-Journal.

Fazit

Am Schluß bleibt nur ungläubiges Staunen, in welcher Geschwindigkeit Veränderungen möglich sind. Mit der konsequenten Anwendung von Dokumentvorlagen in Word und Stilvorlagen in QuarkXPress konnte dem Projekt ein stabiles Fundament gegeben werden. Die Möglichkeiten der Trennung zwischen Inhalt, Struktur und Layout konnten so in anschaulicher Weise zeigen, das erst damit Dokumente schnell und flexibel für einzelne Anwendungen mühelos transformiert werden können. Die direkte Verwendung von XML zeigte sich als sehr großer Vorteil, da mehrfache Erfassungen von Texten nun der Vergangenheit angehören. Neben diesen Vorteilen ist aber auch einiges zu beachten. Für die Umstellung bedarf es Mühe, Zeit und Überzeugungskraft vor allem bei den Autoren; sie liefern die vorbereiteten Daten. Es hilft nicht den Versprechungen zu glauben, das in einem Speichern im XML-Format der Stein der Weisen liegt. Erst in der Kombination von Formatvorlagen und Tags die diese Informationen weitertragen ist eine Lösung zu sehen, da nur der Autor seinen Inhalt mit Formatvorlagen explizit und in der Struktur erkennbar machen kann.

Und was bleibt noch zu sagen: Diese Lösung passt nicht nur für das *cms-journal*, sondern kann für andere Zeitschriften und Buchprojekte angewendet werden.

Quellen

[1] Hinweise für Autoren (Word-Vorlage zum Herunterladen, Benutzungsanleitung etc.) http://edoc.hu-berlin.de/e_rzm/hinweise.php

[2] Dokumenten- und Publikationsserver der HU *http://edoc.hu-berlin.de/e_rzm*

Personalia

Herr Tobias Buddensieg hat am 1.6.2002 seine Tätigkeit in der Abteilung Multimediaservice als Vertretung für Frau Scholz (Elternzeit) aufgenommen.

Herr Roland Herbst ist seit dem 1.7.2002 in der Abteillung DV in der Verwaltung für Netzwerk- und Serversicherheit verantwortlich.

Herr Daniel Rohde arbeitet seit dem 1.1.2003 in der Abteilung Systemsoftware und Kommunikation.
Herr Sascha Demon hat seine Tätigkeit am 1.1.2003 im PC-Saal aufgenommen.

Wir wünschen allen viel Freude und Erfolg bei ihrer Tätigkeit.

Frau Friderike Schimmelpfennig hat zum 30.6.2002 ihre befristete Tätigkeit im RZ beendet.
Wir wünschen Frau Schimmelpfennig für die Zukunft alles Gute.