

XML IN DE GRAFISCHE INDUSTRIE

XML (eXtensible Markup Language) zal als standaard ook binnen de grafische industrie een steeds grotere rol gaan spelen. Op XML gebaseerde standaards zoals JDF (Job Definition Format), SVG (Scalable Vector Graphics) en ebXML (electronic business XML) worden op dit moment al toegepast in grafische systemen en software.

Maar ook opdrachtgevers die gebruik (gaan) maken van op XML gebaseerde content management systemen zullen naar verwachting de uitvoer daarvan ook verder willen laten verwerken tot druk- en printwerk.

Klanten beschikken nu of op termijn over systemen die XML uitvoer kunnen genereren: databases, content management systemen, websites maar ook de nieuwste versies van desktop software zoals Microsoft Excel en Access.

De grafische serviceverlener kan hierop inspelen door een probleemloze koppeling tot stand te brengen tussen deze XML informatiebronnen en opmaaksoftware zoals Quark XPress, Adobe InDesign en Adobe FrameMaker. Ook kunnen conversiediensten worden aangeboden om XML vanuit de opmaak terug te leveren als XML of kan XML worden gebruikt bij het bouwen van websites voor klanten.

Het is dus belangrijk om te weten hoe dit kan, en op welke zaken dan gelet moet worden.

ASVAT revisited ?

Oudgedienden in de grafische industrie herinneren zich wellicht de vanuit het KVGGO ondernomen poging om orde te scheppen in de toentertijd (midden jaren '80) ontstane chaos rondom het 'gecodeerd' aanleveren van tekst vanuit tekstverwerkers. Dit leidde tot een coderingsinstructie die de naam ASVAT kreeg. Daarin stond precies aangegeven hoe de klant een tekst in de tekstverwerker kon voorzien van codes die, na conversie door een zogenaamde 'floppy kraker', automatisch strokenzetsel opleverde met de juiste typografische specificaties. Dit vergde wel eenmalig afstemming met een klant en aanmaak van conversietabellen en codemacro's (nu zouden we dit typogrammen noemen), maar daarna kon tijd en geld worden bespaard bij herhalingsopdrachten.

De praktijk was echter weerbarstig; klanten maakten wel eens een typefout in een code, of gingen zelf codes bij verzinnen die niet vooraf afgesproken waren. Of men vergat de code voor vet zetten weer uit te zetten, en dat leidde dan soms tot lange lappen strokenzetsel die direct de afvalbak in konden.

Helpt XML ?

Als XML indertijd al beschikbaar zou zijn geweest dan had men deze situatie het hoofd kunnen bieden. Ten eerste is XML niets anders dan een standaard voor het coderen van informatie die hiërarchisch gestructureerd is. Alles wat men kan visualiseren in een boomstructuur, van een automotor tot een stamboom, kan in XML gecodeerd worden. En dus ook de inhoud van een boek, want ook dat bestaat uit hoofdstukken, secties, paragrafen, woorden, letters, tabellen en afbeeldingen.

De X in XML staat voor eXtensible, uitbreidbaar. Dat betekent dat een ieder vrij is in het benoemen van zijn eigen codes, ofwel 'tags' in XML jargon. Dat is al een verbetering ten

opzichte van het ASVAT concept, want klanten laten zich nu eenmaal niet makkelijk dwingen in een door anderen bedacht keurslijf.

Een tweede eigenschap van XML documenten is dat het zeer eenvoudig is om te controleren of alle codes netjes worden afgesloten. Daartoe opent men het XML bestand in Internet Explorer (5.0 of hoger) en zal deze precies gaan aangeven of en waar er gezondigd is tegen de coderingsafspraken. In meer technische termen: IE controleert of het XML bestand '*well-formed*' is.

Een derde eigenschap van XML is dat het mogelijk is om de boomstructuur digitaal vast te leggen in een DTD of Schema. Het concept van een DTD (Document Type Definition) is overgenomen van de SGML standaard, die eigenlijk model heeft gestaan bij de ontwikkeling van XML. Bij het maken van een XML document kan men aangeven binnen welke boomstructuur deze moet passen; deze moet bijvoorbeeld gebaseerd zijn op de 'artikel DTD' of de 'nieuwsbrief DTD' die binnen de klantorganisatie is opgesteld. Vervolgens kan worden gecontroleerd of het XML document aan de structuur van de DTD voldoet. Dit gebeurt door een software programma, ook wel een 'parser' genoemd. Deze controle kan men al bij aanmaak van het XML document doen, zodat de auteur zelf de afwijkingen van de structuur kan oplossen. Internet Explorer is niet in staat om deze controle goed te doen, maar er zijn inmiddels veel van dergelijke 'parsers' beschikbaar, of als Open Source software of ingebouwd in een XML editor. Op deze manier kan men als het ware een 'preflight' uitvoeren op het aangeleverde XML document; zitten daar niet afgesproken codes in dan is het XML document niet '*valid*'.

De adder onder het gras is wel dat in de XML standaard het gebruik van een DTD of Schema (een modernere variant van een DTD) *facultatief* is; het mag, maar hoeft niet. Zonder gebruik van een DTD blijft het dus mogelijk dat er opeens nieuwe codes in de XML bestanden opduiken. Ook is er grote kans dat de structuur van dergelijke XML bestanden onhandig of ronduit onbruikbaar is voor geautomatiseerde verwerking.

Aanmaak van XML bestanden

XML bestanden kunnen zoals gezegd worden aangemaakt in een content management systeem. Als een klant dat intensief gebruikt zal er waarschijnlijk ook een DTD voor zijn specifieke documenttypen ontwikkeld zijn, anders helpt het systeem hem niet bij het snel en gestructureerd publiceren van de informatie op de website.

Het is ook mogelijk om XML aan te maken vanuit een echte XML editor zoals XMetaL van Blast Radius, Oxygen, EPIC van ArborText of XML Spy van Altova. Er is zelfs een gratis XML Notepad te downloaden vanaf de Microsoft website. Ook hier is een DTD of Schema zinvol, want deze kan de gebruiker sturen bij het correct toepassen van de structuurafspraken. Bij het werken met externe auteurs zal het wat moeilijker zijn om iedereen met andere software te laten werken. Dan kan software zoals S4/Text van i4i, WordShadow van Ovidius of WorX van HyperVision Ltd. uitkomst bieden. Dit zijn extensies die van MS Word een (bijna) echte XML editor kunnen maken, waarbij de DTD structuur de gebruiker in de juiste richting stuurt via dummy teksten die men zelf kan invullen of via een pop-up menu dat steeds de tags weergeeft die op de cursorpositie zijn toegestaan. Het is zelfs ook mogelijk om allerlei Word functies die tot niet gestructureerd werken zouden kunnen leiden uit te schakelen. Een andere mogelijkheid is het gebruik van Word 2003 Professional, waarmee opslaan als XML volgens een eigen Schema mogelijk is.

Een laatste optie is om via een gespecialiseerde website een on-line XML editor te gebruiken of een conversiedienst aan te spreken die een Word document kan omzetten naar een XML bestand. Met name dit laatste lukt niet volledig automatisch en vergt dus de tussenkomst van een operator.

Van XML bestand naar opmaak

Er zijn verschillende mogelijkheden om een XML bestand automatisch in een opmaakprogramma in te laten lopen met de juiste typografische instellingen.

Men kan gebruik maken van speciaal voor dit doel ontwikkelde opmaaksoftware zoals 3B2 van Advent of XPP van XyEnterprise. Ook Adobe FrameMaker 7.0 kan met XML bestanden werken en zelfs ook de correcties in de pagina terugvoeren naar het XML brondocument.

De nieuwste versies van Quark XPress en Adobe InDesign ondersteunen ook XML import en export, zij het minder geautomatiseerd. Verder zijn er diverse bedrijven zoals Independent System Integrators, TechnoDesign en WoodWing die volautomatische omzetting van XML naar Quark XPress of Adobe Indesign kunnen realiseren.

In alle gevallen moet eerst het nodige voorwerk worden gedaan, zoals vastleggen van de layout en definitie van alinea- en karakterstijlen in een template.

Van opmaak naar XML

Retour leveren van XML vanuit een opgemaakt document zal bij migratie van een klant richting een content management systeem steeds vaker gevraagd worden. Ook hier is weer belangrijk of er een DTD is waaraan de structuur van het resulterende XML bestand moet voldoen. Tekst kan altijd wel geëxporteerd worden, maar het is beter om de typografische instellingen op een slimme manier om te zetten naar tags in XML. Daarvoor zijn ook Xtensies en plug-ins gemaakt, bijvoorbeeld Atomik voor Quark XPress en Smart XML Export van WoodWing voor InDesign. Hoeveel handwerk daar nog bij te pas komt hangt ook af van de manier waarop de DTP-er bezig is geweest; is er consequent met alinea- en karakterstijlen gewerkt met gelinkte tekstkaders dan valt dit mee, is het een chaos dan kost dit erg veel moeite.

XML dienstverlening

Een aantal grafische bedrijven hebben zich al op dit terrein gespecialiseerd; zij bieden in feite een volledig traject aan, van advies tot implementatie en ontwikkeling van DTD's/Schema's en uitvoering van de opmaak.

Maar ook minder gespecialiseerde bedrijven kunnen hierop inspelen door zich – al of niet samen met kennisleveranciers – in de mogelijkheden van XML te verdiepen. De technologie is laagdrempelig; via het web is ontzettend veel informatie en software beschikbaar, de laatste tijd gelukkig ook voor het Macintosh platform.

TIP: Ingangscontrole van XML documenten

- Open ieder XML document in Internet Explorer 5.0 of hoger; er moet dan een nette boomstructuur zichtbaar zijn, als er foutmeldingen optreden is het document niet in orde. Tevens kan nu een indruk worden verkregen van de kwaliteit van het aangeleverde document. Kijk ook naar de manier waarop letters met accenten en symbolen zijn gecodeerd: dat kan Unicode zijn (met conversie perikelen richting MacIntosh) of netjes volgens de Latin1 standaard (bijv. ë of ë in plaats van de ë).
- Vraag bij klanten altijd of er een DTD is; zo ja, dan moet deze ook meegestuurd worden. XML documenten nu ook door een ‘parser’ laten controleren.
- Is er geen DTD toegepast of is deze niet beschikbaar dan is er geen enkele garantie dat het XML bestand beter verwerkbaar is dan een willekeurig tekstbestand; automatiseren van de opmaak is dan een groot risico omdat er iedere keer iets anders in de XML bestanden kan zitten. Het is bijvoorbeeld mogelijk om een lange lap ASCII tekst te voorzien van een begin- en eindcode en deze op te slaan met extensie .xml. Dit levert wel een well-formed XML bestand op maar heeft geen enkel voordeel ten opzichte van aanlevering van het ASCII bestand zelf. Dit zal ook zichtbaar zijn bij openen in Internet Explorer; de boomstructuur bestaat dan maar uit één element.

ir. R.J. Janssen,
Screens & Pages
R.J.Janssen@SandP.nl