

Custom Defined Schema's

JEAN PAOLI, XML-ARCHITECT BIJ MICROSOFT
EN AUTEUR VAN DE XML-STANDAARD



Wie is Jean Paoli? Zijn naam prijkt op de voorkant van de W3C XML 1.0-specificatie. Hij is een van de auteurs en architecten van de XML-standaard en nog steeds een actief lid van de W3C XML working group. Bij Microsoft is hij onder meer verantwoordelijk voor de implementatie van XML in Microsoft Office Systems. Er wordt beweerd dat Paoli dé persoon is die XML zowel bij Microsoft als voor Microsoft succesvol heeft gemaakt. Redenen genoeg hem op te zoeken.

De XML-specificatie is af, wat valt er nog over te vragen? Het zou een interview kunnen worden waarbij Paoli elke vraag zou kunnen beantwoorden met een verwijzing naar een bepaalde bladzijde van de XML-specificatie. Maar het werd een boeiend gesprek. Nog voordat een vraag kon worden gesteld, stak Paoli van wal. "Voordat ik in mei 1996 bij Microsoft ben begonnen, was ik vooral bezig het gebruik van SGML te stimuleren", vertelt hij. In die tijd was Paoli technical director bij Grif, een bedrijf dat zich voor 100% op SGML richtte. Grif was onder andere leverancier van de SGML-editor voor Windows, Unix en Macintosh. Bovendien was hij aan de Universiteit van Parijs verbonden en hielp hij mee met allerlei researchprojecten voor SGML. "Via de universiteit kwam ik in contact met W3C", zegt hij. "De SGML-wereld was klein, ik kwam vaak dezelfde mensen tegen bijvoorbeeld op de jaarlijkse SGML-conferentie." Twee weken na de SGML-conferentie in 1996 maakte hij zijn overstap naar Microsoft bekend.

Versimpel SGML

Paoli werkte in de Platform Group en was betrokken bij de ontwikkeling van IE40. Bovendien werd hard gewerkt aan 'Trident', de opvolger van de Mosaic html-engine. In die tijd was de browseroorlog in volle gang. Enkele maanden daarvoor, in december 1995, had Bill Gates op 'internet day' de grote internetommezwaai van Microsoft bekend gemaakt. Elk product moest 'internet enabled' worden en html kunnen lezen en schrijven. Maar er moest meer gebeuren. Het was duidelijk dat SGML te complex werd bevonden. Door zijn contacten met de W3C, zat Paoli al gauw voor Microsoft in de W3C SGML Editorial Review Board. Een van de eerste taken was de simplificatie van de SGML-spec. Paoli benadrukt: "Eén van mijn eisen was om SGML makkelijker te maken. Haal alle overbodige zaken eruit. Ik was indertijd in het Internet Explorer 4.0-team belast met databinding en had alle tijd om aan deze nieuwe specificatie mee te werken. Bovendien lag de focus toen voornamelijk op andere onderdelen van de IE, zodat er geen druk op ons lag."

De standaard is af

Op de vraag hoe de XML 1.0-specificatie tot stand is gekomen en hoe dat proces is verlopen, antwoordt Paoli: "We hebben de XML-specificatie in vier à vijf maanden gemaakt met een klein team dat bestond uit Tim Bray, C. M. Sperberg-McQueen en Jon Bosak die bekend zijn vanuit de SGML-wereld. De spec is voor 80% op SGML gebaseerd. XML is 'more strict' dan SGML, waardoor het eenvoudiger is te genereren en te verwerken. Het aantal parsers voor SGML is zeer beperkt, maar voor XML zijn er vele beschikbaar. Op elk platform zijn er bovendien verscheidene parsers beschikbaar, standaard

zitten er al twee in Windows XP. Een ander verschil is dat bij SGML een DTD (document type declaration) vereist is en dit bij XML optioneel is, zodat het eenvoudiger in gebruik is." Inmiddels is de derde revisie van de 'Extensible Markup Language (XML) 1.0 W3C recommendation 04 februari 2004' uit. Paoli vertelt dat er eigenlijk alleen maar kleine uitbreidingen en correcties in de revisie 2 en 3 te vinden zijn. Er zijn in de toekomst geen grote wijzigingen te verwachten. De standaard is af. Wel zijn er volgens Paoli nog veel uitbreidingen en afgeleide standaards te verwachten.

We willen jullie data

De eerste taak was om XML in de browser te implementeren. Paoli: "Bij Microsoft hebben we toen snel kunnen schakelen en wij hebben vrij snel de eerste MSXML opgeleverd. Naast de client was mijn eigen doelstelling om XML zo snel mogelijk op de server beschikbaar te krijgen. Een van de eerste serverapplicaties was BizTalk die met XML overweg kon. Daarna volgde snel SQL Server en andere. En dit is pas het begin, kijk maar wat SQL Server Yukon gaat brengen. Ik weet nog dat ik in een meeting met IBM grapte: 'We willen jullie data'. Het ontsluiten van opgeslagen gegevens, op een mainframe of server moet gewoonweg eenvoudiger. Gelukkig was IBM een van de eerste supporters van de nieuwe XML-standaard en begon actief mee te helpen met afgeleide en aanvullende standaards zoals SOAP. IBM is tegenwoordig ook betrokken bij de verschillende WS-* initiatieven."

Custom Defined Schema's in Office

"Na mijn werk in de platform group heb ik gewerkt in het Office-team", vervolgt Paoli. "Een grote uitdaging was om de Office-familie volledig XML-aware te maken. Het resultaat is in elk Office-onderdeel terug te vinden, maar vooral in een compleet nieuw product InfoPath dat XML kan bewerken. Zowel Word als Excel kunnen nu native in XML hun bestanden opslaan." Op de tegenwerping dat het wordML-formaat en het spreadsheetML-formaat behoorlijk complex is, antwoordt Paoli bevestigend. "Klopt, helaas worden de XML-mogelijkheden in Office nog niet volledig benut door ontwikkelaars. Maar klanten willen archiveren in XML-formaat in plaats van het .DOC-formaat, en dat kan nu. De ware schoonheid en kracht zitten in de mogelijkheid van Custom Defined Schema's. Daar moeten ontwikkelaars gebruik van maken en ik verwacht dat de komende jaren vele schema's gemaakt zullen worden. Er is helaas geen universeel schema, er komen miljoenen schema's. Ontwikkelaars kunnen deze eenvoudig maken met XMLSpy of Visual Studio of zelfs door power users met InfoPath. In bedrijven en zelfs in afdelingen zullen eigen schema's komen voor de uitwisseling van data. Dat is veel efficiënter dan het rondsturen van documenten. De volgende stap is dus informatie-uitwisseling via XML met Custom Defined Schema's."