

Dana Florescu (W3C) over XML & XQuery:

“Rijker dan het relationele model”

Peter Verkooijen

XQuery 1.0 nadert eindelijk voltooiing. Een W3C-werkgroep sleutelt al bijna zeven jaar aan de querytaal voor XML. Is het wel een goed idee om XML in te zetten voor datamanagement?

Dana Florescu zegt dat het World Wide Web Consortium (W3C) “binnen een paar maanden” XQuery 1.0 als officiële aanbeveling zal publiceren. Florescu is lid van de werkgroep die sinds september 1999 aan de querytaal voor XML werkt. “Het proces hangt af van factoren die niet te voorspellen zijn”, zegt Florescu. Het W3C heeft voor de kandidaat-standaard een testpakket met twaalfduizend tests voorgeschreven. “Voor elke functie in de taal moeten we bewijzen dat er tenminste twee interoperabele implementaties zijn, anders blijven ze niet in de taal. Het W3C gaat zeer grondig te werk.”

Eigen querytaal

Het idee dat XML zijn eigen querytaal nodig had ontstond eind jaren negentig. Een serie andere talen ging XQuery voor. Jonathan Robie was een van de ontwikkelaars van XQL (XML Query Language), een aanvulling op Extensible Stylesheet Language (XSL). Daaruit ontstond XPath waarmee data uit XML documenten zijn te halen. In 1999 vormde het W3C een werkgroep om een volwaardige querytaal voor XML te ontwikkelen. Naast Robie en Florescu nam daarin Don Chamberlin van IBM plaats. Chamberlin stond eerder aan de wieg van SQL. Florescu vertegenwoordigt in de werkgroep Oracle en werkte eerder aan OQL (Object Query Language).

XQuery bouwt op drijfzand en lijkt volgens Date meer op een programmeertaal

De werkgroep stelde een lijst met negen voorbeelden van toepassingen samen, waaronder toegang tot data in een relationele database en data in SGML (Standard Generalized Markup

Language) documenten. Uit deze cases destilleerde de groep negentig query's die de nieuwe taal moest kunnen uitvoeren. Jonathan Robie heeft gezegd dat XQuery vooral het programmeren met XML gemakkelijker moest maken. Een van de innovaties in XQuery ten opzichte van XPath en XSLT (XSL Transformations) zijn de FLWOR expressies. FLWOR staat voor “For, Let, Where, Order by, Return”. Dankzij FLWOR kan XQuery in XML documenten niet alleen data vinden, maar ook correleren en nuttige resultaten produceren.

De belangstelling voor een querytaal voor XML nam toe door de groei in webdiensten. XML werd het esperanto voor handel op het internet. Hoewel XML steeds vaker wordt gebruikt voor data-uitwisseling, blijft de relationele database de standaard voor dataopslag. Toch is dit niet het dilemma dat de XQuery-werkgroep specifiek voor ogen had. “XML is niet voor één enkele gebruikersgroep”, zegt Dana Florescu. “Dat is het voordeel van XML. Mensen gebruiken XML op verschillende manieren voor verschillende soorten informatie, van gestructureerde data zoals in relationele databases tot documenten in content management systemen en Word-bestanden, Powerpoint-presentaties en spreadsheets. Hetzelfde datamodel kan verschillende vormen van informatie representeren. XQuery is aantrekkelijk voor alle XML gebruikers.”

Heb je XQuery wel nodig als je alle data consequent in een relationele database onderbrengt? Dana Florescu vindt dat geen realistisch scenario. “Voor een relationele database moet je eerst een schema ontwerpen”, zegt ze. “Niet in alle gevallen is vooraf duidelijk wat het schema is. Als de applicatie en data zich ontwikkelen heb je misschien andere velden nodig. Het systeem moet uitbreidbaar zijn. In een relationeel systeem moet je alle tabellen en columns vantevoren specificeren. Steeds meer mensen gebruiken een mix van gestructureerde data in een relationele database en ongestructureerde data in XML.”

Stap terug

De term 'ongestructureerde data' werkt als een rode lap op de verdedigers van het relationele model. Ongestructureerde data zijn volgens hen data die nog niet in een database zijn ondergebracht. En eigenlijk zijn ongestructureerde data helemaal geen data, maar ruis of troep. Ongestructureerde data mogen dan gemakkelijk in XML documenten zijn onder te brengen, maar dat maakt ze nauwelijks hanteerbaarder.

Chris Date, pionier en hoeder van het zuivere relationele model, ziet in XML een poging het achterhaalde hiërarchische datamodel opnieuw uit te vinden. XQuery bouwt op drijfzand en lijkt volgens Date meer op een programmeertaal dan een querytaal. "Als je SQL goed vond, zul je XQuery helemaal te gek vinden", zegt Date. En dat is geen aanbeveling.

Chris Date staat niet alleen in die opvatting. "XML en XQuery zijn een stap terug naar het hiërarchische model van dertig jaar geleden", zei Bryn Rhodes van Dataphor eerder dit jaar in Database Magazine. "Waarden zijn niet altijd hetzelfde. Het relationele model was juist ontworpen om die problemen te elimineren. Ik kan XML als markup-taal nog accepteren. Om het uit te breiden naar een algemeen datamodel is een stap terug. En de querytaal is een gruwel. XML verdwijnt, het is een rage die vanzelf zal wegebben."

Glenn Reid probeert met zijn website XMLsucks.com zijn steentje bij te dragen aan de ondergang van XML. Reid heeft sinds midden jaren tachtig als software-ontwikkelaar voor Adobe en Apple gewerkt. XML kwam voort uit die desktop publishing-wereld, niet uit datamanagement. SGML en later HTML zetten markeringen in tekst die aangaven hoe de parser of browser de tekst moest weergeven. "XML keert dat binnenstebuiten", zegt Reid. "Je kunt een leeg XML document hebben vol haakjes en datadefinities, maar zonder data. Men noemt XML een markup-taal omdat het de syntax van HTML leent, maar eigenlijk is XML een *container file format*."

XML verbindt geen semantische betekenis aan de markeringen. Applicaties moeten onderling afspreken hoe ze de labels definiëren. "Het feit dat XML geen semantische informatie bevat betekent dat het niks meer doet dan de data zelf representeren", zegt Reid. "En niet op de meest efficiënte manier. Wat je in een SQL database opslaat mag alle karakters bevatten die je nodig hebt. Binnen XML mag je geen haakjes en een heleboel andere karakters gebruiken. Je moet data aanpassen om het in XML op te slaan."

Is XML dan helemaal nergens goed voor? "XML is bruikbaar om data in kale ASCII te vertalen voor verzending", zegt Reid. "Dat was de oorspronkelijke functie. XML is op zichzelf niet slecht, maar wordt te vaak gebruikt waar het niet hoort. Het doet niets voor je. Het presenteert de data en daar zijn genoeg andere methoden voor." XML voegt volgens Reid weinig toe aan wat al kon met *comma separated values* (CSV). CSV is niet goed bruikbaar voor hiërarchische data op verschillende niveaus, maar een hele reeks alternatieven zoals YAML, OGD, JSON en

SSYN zou daar oplossingen voor hebben zonder de onhandige labels van XML.

Niet onnodig gecompliceerd

De labels in XML zijn het resultaat van het uitgangspunt dat zowel mensen als machines de bestanden moesten kunnen lezen. Dat heeft waarschijnlijk veel bijgedragen aan de populariteit van XML, maar XML bestanden zijn daardoor veel zwaarder dan ze zouden hoeven zijn. XML bestanden vergen vijf tot tien keer meer processorkracht dan applicatiespecifieke binaire bestanden.

"XML is breedsprakig, maar dat is geen probleem", stelde het Nederlandse W3C-lid Bert Bos in een beruchte opsomming van XML in tien punten uit 1999. "Het was een bewuste keuze van de XML-ontwikkelaars." Compressie en dalende prijzen voor opslag zouden de extra last compenseren volgens Bos, maar met de groeiende populariteit van XML groeit de last exponentieel. Er is een markt ontstaan voor hardware XML versnellers. Links en rechts zijn lichte varianten van XML voor specifieke toepassingen opgedoken. Het W3C heeft aan een standaard voor een binaire versie van XML gewerkt, tot dusver zonder concreet resultaat. Of het een goed idee was op dit fundament een datamanagement-systeem te bouwen is nooit besproken in de XQuery-werkgroep.

SQL-fans zullen gek zijn op XQuery

"We wilden een querytaal voor XML data hebben, dus het is vanzelfsprekend op XML gebaseerd", zegt Dana Florescu. "Wat is het alternatief? Veel mensen bekritisieren XML, maar iedereen gebruikt het. Ik zou XML ook niet hebben ontworpen zoals het nu is, maar de critici missen het historisch perspectief. XML is gebaseerd op dertig jaar werk in SGML, waar HTML uit voortkwam. XML was een natuurlijk uitvloeisel. Het was ook geen goed idee geweest als het W3C HTML zou hebben genegeerd. XML is niet simpel, maar ik geloof niet dat het onnodig gecompliceerd is. Het probleem dat het probeert op te lossen, al die verschillende vormen van informatie in een simpel model weergeven, is inherent complex."

Maar lost XML dit probleem werkelijk op? Chris Date's Incoherentie Principe is van toepassing: "Het is niet mogelijk iets dat incoherent is coherent te behandelen." XQuery probeert het datamanagement-wiel uit te vinden met een vierkant wiel, volgens Date's handlanger Fabian Pascal. XML heeft oppervlakkig gezien kenmerken van een DBMS, zoals opslag, schema's en querytalen, maar mist elementen als indexen, beveiliging, transactie- en data-integriteit, toegang voor meerdere gebruikers, triggers enzovoort.

Glenn Reid kan zich weinig voorstellen bij een 'XML database'. "XML bevat niet de semantiek van de data", zegt Reid. "Het DTD-bestand moet aangeven wat de velden betekenen. Dat is het

grappige van XML. Alle interessante, bruikbare elementen zitten om XML heen. Je hebt nog steeds een database-model bovenop XML nodig. Toch wordt XQuery gepresenteerd als vervanger van SQL. Er kan nog een relationele database onder zitten, maar het is gericht op mensen die hun data in XML opslaan. XML is niet erg sterk als uiteindelijk database-vorm. Dat vind ik het verwarrende aan Xquery; welk probleem probeert het op te lossen?"

XML in relationele databases

XML was oorspronkelijk bedoeld voor uitwisseling van data, niet voor datamanagement. Dana Florescu ziet dat anders. "Ik geloof niet dat de mensen die XML hebben ontwikkeld, zoals Tim Bray en Michael Sperberg-McQueen, ooit hebben gezegd dat XML niet voor datamanagement was", zegt Florescu. "Het was een model voor weergave en uitwisseling van data en zo wordt het tegenwoordig toegepast. Data worden steeds vaker niet alleen uitgewisseld in XML, ze blijven in XML. Nee, dat is niet gevaarlijk. Dat is erg positief. Het heeft veel voordelen. Het is een rijker model dan het relationele model."

Ongestructureerde data uit XML halen en in de structuur van relationele tabellen onderbrengen is volgens Dana Florescu aanvankelijk wel geprobeerd. Dat heette *shredding*. Data moesten worden versnipperd over de tabellen. "Het werkt niet erg goed", zegt Florescu. "Je verliest veel informatie. De rangorde van een

XML document doet er toe. Die informatie behouden als je het overzet naar tabellen vereist veel werk." Oracle en IBM ontwikkelden volgens Florescu methoden om XML diep binnen relationele databases op te slaan. Om die data te benaderen moet je SQL/XML gebruiken, een hybride querytaal van SQL met XQuery extensies. Gaat XQuery SQL verdringen? "Zo zou ik het niet typeren", stelt

Veel mensen bekritisieren XML maar iedereen gebruikt het

Florescu. "Ze vullen elkaar aan. SQL vervangen was nooit het doel. Er is een grijs gebied waar ze elkaar overlappen, maar het is geen erg groot gebied. SQL is erg goed in gestructureerde data, zoals bankrekeningen. XML is goed voor irreguliere data, zoals Word-bestanden. Elk heeft zijn eigen functie. Natuurlijk is er overlap, maar ik geloof niet dat XQuery ooit SQL zal vervangen in een bank of financiële instelling. XML is een nieuwere technologie. Het zal jaren, misschien decennia duren voor XML technologieën net zo goed presteren als relationele technologieën die al 25 jaar worden geoptimaliseerd."

Peter Verkooijen is freelance journalist.

Update

Sybase breidt Risc Analytics Platform uit met Applix TM1

Sybase heeft een overeenkomst gesloten met Applix voor het in licentie leveren van de TM1 analysesoftware. Op basis van deze overeenkomst gaat Sybase de 64-bit versie van TM1 bundelen met het Sybase Risc Analytics Platform (RAP) voor real-time risicoanalyses van financiële transacties, en andere risico-management-toepassingen. Door de nieuwe samenwerking kunnen klanten de analysesnelheid van TM1 op het 64-bit platform combineren met Sybase's Risc Analytics Platform voor het real-time managen en analyseren van steeds grotere datahoeveelheden.

Door de toenemende datahoeveelheden die bij financiële toepassingen worden gegenereerd, en verdere ontwikkeling van kwantitatieve analysemethoden, staat de analyse-infrastructuur van handelshuizen en grote bedrijven onder druk. Cross-asset handel, complexe hedging-

technieken en de behoefte aan risico-management in elke procesfase, verhogen de eisen die worden gesteld aan het real-time analyseren van zowel live data als historische dataverzamelingen. Ook nieuwe richtlijnen, zoals de Regulation NMS (National Market System) en MiFID (Markets in Financial Instruments Directive) dwingen organisaties tot real-time rapporteren en het genereren van auditoverzichten tijdens het gehele proces. Applix TM1 is een performance management-applicatie voor strategische business planning, budgettering, rapportages en 'what if'-analyses van grote hoeveelheden operationele data, verkoopresultaten, personeelsinformatie en andere bedrijfstoepassingen. De software kreeg bij de OLAP Survey de hoogste scores voor querysnelheid, implementatie, datacalculaties, totale bedrijfsvoordelen en het minste aantal technische problemen.

Zie www.applix.com en www.sybase.nl

SAS-oplossing verbetert marketingprestaties

SAS maakt met nieuwe software inzichtelijk wat de bijdrage van marketingactiviteiten aan de inkomsten is. SAS for Marketing Performance Management verbetert de marketingprestaties en kwantificeert de marketingresultaten, en laat ook duidelijk zien wat de invloed van marketing is op de prestaties van het bedrijf. Het is geïntegreerd in SAS' customer intelligence-oplossingen en zorgt voor de onderlinge afstemming van marketingstrategieën en -activiteiten, het koppelt marketingprestaties aan de financiële prestaties van het bedrijf en zorgt ervoor dat marketingactiviteiten kunnen worden gekwantificeerd.

In een onderzoek van 15 augustus 2006 concludeert onderzoeksbureau Forrester dat SAS met SAS Marketing for Performance Management ver vooruit loopt op haar concurrenten.

Zie www.sas.com