

Query-taal voor XML wordt maar langzaam volwassen

XQuery zet zijn eerste schuchtere stappen

Hans C. Arents

De XML-querytaal XQuery is na meer dan twee jaar van ontwikkeling en overleg nog steeds niet vastgelegd in een definitieve standaard. Erger nog, zelfs in de meest recente werkversie van de standaard wordt nog niets gezegd over mogelijke functionaliteiten voor 'insert', 'update' en 'delete' - cruciaal om XQuery te laten uitgroeien tot een volwaardige taal voor datamanipulatie. Wie toch al wat ervaring met XQuery wil opdoen, kan intussen terecht bij een aantal XML-database-leveranciers, die -met succes- een proefversie van XQuery hebben geïmplementeerd.

In een vorig artikel, *Voorproefje XQuery stemt gematigd optimistisch* (zie Database Magazine, nummer 2, april 2001) hebben we besproken hoe het World Wide Web Consortium (W3C), de non-profitorganisatie die instaat voor het vastleggen van alle essentiële webstandaarden, er eindelijk in geslaagd was een query-datamodel en een query-algebra voor XQuery te definiëren, en begonnen was met het vastleggen van de syntax voor XQuery. Bijna een jaar later, op 20 december 2001, is een nieuwe werkversie van de Xquery-standaard verschenen. Deze werkversie verschilt niet wezenlijk van de vorige werkversie van de Xquery-standaard, op één belangrijk aspect na: XQuery maakt nu gebruik van XPath 2.0. Dit is de nieuwste versie van XPath, de zoeksyntax die het mogelijk maakt bepaalde knoppen in een XML-boomstructuur terug te vinden en te selec-

teren. XPath 2.0 speelt ook een sleutelrol in de XML-stylesheet-taal XSLT. Daarvan is tegelijkertijd met de nieuwe werkversie van XQuery een eerste werkdocument voor de nieuwste versie 2.0 verschenen.

De XML Query Working Group en de XSL Working Group werken momenteel dan ook nauw samen om ervoor te zorgen dat hun inspanningen op een gelijksoortige manier gebruik maken van de XPath 2.0-syntax. De proliferatie aan nieuwe, goedgekeurde of nog in ontwikkeling zijnde XML-standaards in het afgelopen jaar heeft er immers voor gezorgd dat momenteel dringend behoefte bestaat aan coördinatie en samenwerking tussen de verschillende XML-werkgroepen van het W3C.

KRITIEK

De nieuwe werkversie van XQuery heeft echter al heel wat kritiek losgemaakt. Had XPath vroeger een vrij eenvoudige syntax en semantiek, XPath 2.0 evolueert meer en meer in de richting van een echte programmeertaal. Dat maakt het heel wat

moeilijker om haar aan te leren en te gebruiken. De voornaamste kritiek is echter gericht op de Xquery-taal zelf, omdat deze nog steeds niet de nodige functiona-

Toch valt met een aantal XML-databaseproducten ervaring op te doen

liteiten bevat om XML-elementen aan te maken, aan te passen of te verwijderen.

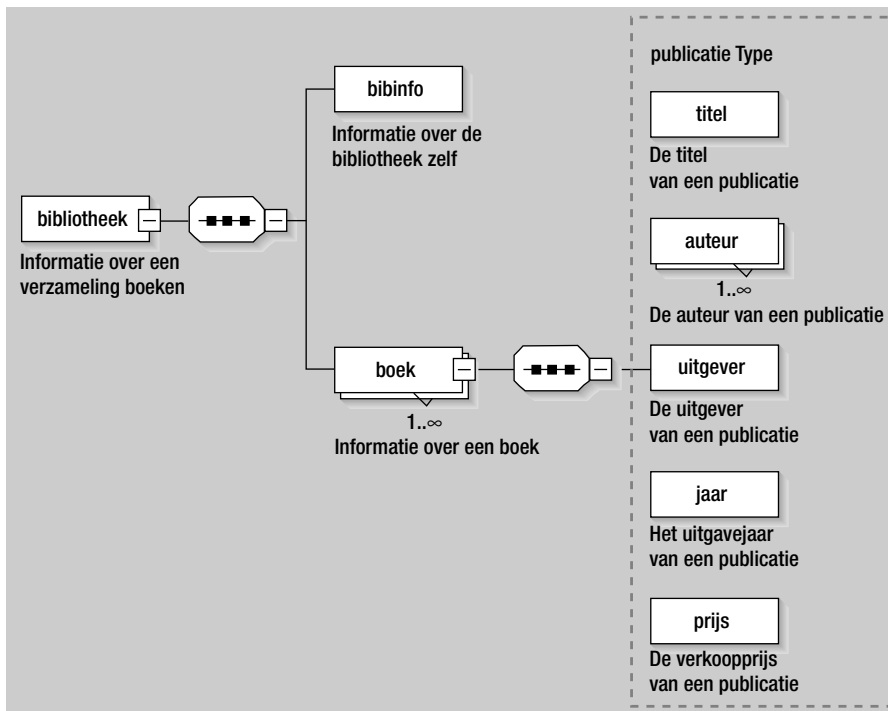
Het W3C verdedigt zich tegen deze kritiek door erop te wijzen dat zijn prioriteiten nu liggen bij het zo snel mogelijk vastleggen van een stabiele standaard XQuery 1.0. De verlangde insert/update/delete-functionaliteiten zouden pas aan bod kunnen komen in een versie 2.0 van XQuery.

XQUERY-UPDATES

Op de recente conferentie XML 2001 in Orlando heeft Jonathan Robie, een van de grondleggers van de huidige versie van XQuery, toch al een eerste voorstel gedaan

Symbol	Betekenis
.	de inhoud van de knop zelf
/	de wortelknop, of de separator tussen de verschillende stappen in het zoekpad
//	de afstammelingen van de knop (op een willekeurige diepte in de boomstructuur)
[...]	een Boleaanse expressie om te filteren in de resultaatknoppen van een bepaalde stap

TABEL 1: OVERZICHT GEBRUIKTE XPATH-SYMBOLEN IN XQUERY-UPDATES.



FIGUUR 1: GRAFISCHE VOORSTELLING VAN HET DOCUMENTMODEL VOOR HET VOORBEELDDOCUMENT "BIB.XML" (GEBRUIKMAKEND VAN XML SCHEMA-SYNTAX).

("Updates in Xquery") voor de mogelijke implementatie van insert/update/delete-functionaliteiten in XQuery.

Om een idee te geven van hoe Xquery-updates eruit zouden kunnen zien, volgen hieronder enkele (eenvoudige) voorbeelden. We gaan ervan uit dat de lezer reeds vertrouwd is met XQuery en de door Xquery gebruikte Xpath-zoeksyntax (zie het vorige artikel). Tabel 1 toont de Xpath-symbolen die we in onze Xquery-updates gebruiken en hun betekenis.

Net als in het vorige artikel werken we met een XML-document met als naam "bib.xml" dat een lijst van <boek>-elementen bevat. Ieder boek-element bevat op zijn beurt een element <titel>, één of meer <auteur>-elementen, en de elementen <uitgever>, <jaar> en <prijs>. Een grafische voorstelling van het documentmodel (in XML Schema-syntax) voor dit "bib.xml"-document geeft figuur 1.

Toevoegen (insert)

Met de INSERT-instructie kan een element worden toegevoegd vóór of na een bestaand element.

Voorbeeld: voeg een nieuw <auteur>-element toe dat de waarde "An Janssens" heeft, in alle boeken die een element

<auteur> bevatten met de waarde "Jan Janssens". Deze nieuwe auteur An Janssens dient te worden toegevoegd vóór de bestaande auteur Jan Janssens.

```
UPDATE
INSERT <auteur>An
Janssens</auteur>
PRECEDING //auteur[.="Jan
Janssens" ]
```

Voorbeeld: voeg een nieuw auteur-element toe dat de waarde "An Janssens" heeft, in alle boeken die een element auteur bevatten met de waarde "Jan Janssens". Deze nieuwe auteur An Janssens moet worden toegevoegd na de bestaande auteur Jan Janssens.

```
UPDATE
INSERT <auteur>An
Janssens</auteur>
FOLLOWING //auteur[.="Jan
Janssens" ]
```

De INSERT-instructie dient niet alleen om een element toe te voegen vóór of na een bestaand element, maar is ook inzetbaar om een attribuut toe te voegen aan een bestaand element.

Voorbeeld: voeg een attribuut aanspreektitel toe dat de waarde "Dr." heeft, in alle <auteur>-elementen met de waarde "Jan Janssens".

```
UPDATE
INSERT ATTRIBUTE aanspreektitel { "Dr." }
INTO //auteur[.="Jan Janssens" ]
```

Met de INSERT INTO-instructie kan men ook een element uitbreiden met een ander element. Dit element wordt dan toegevoegd aan het einde van de lijst met kind-elementen van het geselecteerde element.

Voorbeeld: voeg een nieuw element <prijs> toe dat de waarde "16" heeft, in alle boeken die een <auteur>-element bevatten met de waarde "Jan Janssens". Dit nieuwe <prijs>-element wordt toegevoegd als het laatste element van de kind-elementen van het <boek>-element.

```
UPDATE
INSERT <prijs>16<prijs>
INTO //boek[auteur="Jan Janssens" ]
Aanpassen (update)
```

De REPLACE-instructie is bedoeld om een element in zijn geheel te vervangen.

Voorbeeld: vervang in alle boeken die een <auteur>-element bevatten met de waarde "Jan Janssens" dit element <auteur> door een nieuw, met de waarde "An Janssens".

```
UPDATE
REPLACE //auteur[.="Jan Janssens" ]
WITH <auteur>An Janssens</auteur>
```

Verwijderen (delete)

Met de DELETE-instructie kan een element worden verwijderd.

Voorbeeld: verwijder alle boeken die een <auteur>-element bevatten met de waarde "Jan Janssens".

```
UPDATE
DELETE //boek[auteur="Jan
Janssens" ]
```

Gebruik binnen een FLWR-expressie

De instructies INSERT, REPLACE en DELETE zijn te combineren binnen een XQuery FLWR-expressie, een expressie opgebouwd aan de hand van FOR-, LET-, WHERE- en RETURN-clauses. Zo kunnen meerdere instructies toevoegen/aanpassen/verwijderen samen worden uitgevoerd.

Voorbeeld: doorloop alle boeken, kijk of het boek een element <auteur> bevat met de waarde "Jan Janssens", en als dit het geval is, voeg een nieuw <auteur>-element met de waarde "An Janssens" toe vóór dit element, en vervang het <auteur>-element met de waarde "Jan Janssens" door een <auteur>-element met de waarde "Jan Pieters".

```
UPDATE
FOR $b IN
document("bib.xml")//boek
LET $a := $b/auteur[.="Jan
Janssens"]
WHERE not (empty($a))
INSERT <auteur>An
Janssens</auteur>
PRECEDING $a
REPLACE $a
WITH <auteur>Jan Pieters/au-
teur
```

PRODUCTEN MET XQUERY

Bovenstaande voorbeelden geven slechts een eerste voorproefje van hoe Xquery-updates eruit kunnen gaan zien. Wie liever zelf wil uitproberen hoe XQuery in de praktijk aanvoelt, heeft inmiddels de keuze tussen een aantal XML-databaseproducten waarvoor een proof-of-concept Xquery-processor geïmplementeerd werd.

De bekendste is momenteel QuiP van Software AG, de makers van de goed verkopende Tamino XML-database. QuiP is een prototype Windows-visuele Xquery-builder, waarmee u Xquery-query's kunt formuleren en de resultaten daarvan meteen kunt bekijken. De software komt met 76 voorbeeld query's, en een aantal XML-bestanden waarop u die query's kunt loslaten. Gebruikers van Tamino kunnen vanuit QuiP ook de XML-database

benaderen. Ondersteuning van de huidige versie van de Xquery-standaard is vrij volledig, en de meest recente versie van QuiP ondersteunt ook het "Updates in XML"-voorstel.

Een ander uitstekend prototype Xquery-processor is de Xquery-demo van

Helaas is XQuery snel aan het uitgroeien tot een taal die veel complexer en minder functioneel is dan SQL

X-Hive Corporation. Daarmee brengt het van de database X-Hive/DB bekende Nederlandse bedrijf één van de meest geavanceerde native XML-databaseproducten op de markt. X-Hives Xquery-demo draait rechtstreeks bovenop X-Hive/DB, maar is daarbij alleen via een

online webinterface toegankelijk. De demo ondersteunt het merendeel van de XQuery Use Cases uit de Xquery-standaard, en u kunt deze query's uitproberen op voorbeeld XML-documenten in de database. U kunt ook eigen query's formuleren en uitproberen. Hoewel Xquery-demo nog geen updates ondersteunt, is dat wel het geval met Xquery-query's die over meerdere documenten heen lopen, zoals vastgelegd in de experimentele XMach-1 XML-database-benchmarking-specificatie.

Microsoft heeft een vergelijkbare Xquery-demo; via een online webinterface zijn enkele voorbeeld Xquery-query's te beoefenen. Ook hier vinden we al voorbeelden van Xquery-updates, maar wel in een syntax die enigszins verschilt van de voorgestelde syntax. Big Green heeft al laten doorschemeren dat hij in de volgende versie van zijn rdms SQL Server

Informatie op Internet

www.w3.org/XML/Activity#query-wg (XML Query Working Group)
www.w3.org/Style/XSL/ (XSL Working Group)
www.w3.org/TR/query-datamodel/ (XQuery 1.0 and XPath 2.0 Data Model)
www.w3.org/TR/query-algebra/ (XQuery 1.0 Formal Semantics)
www.w3.org/TR/xquery (de XQuery 1.0 taal)
www.lehti.de/beruf/diplomarbeit.pdf (thesis waarop "Updates in Xquery"-voorstel gebaseerd is)
<http://dbs.uni-leipzig.de/en/projekte/XML/XmlBenchmarking.html> (XMach-1 XML-database-benchmark)
www.brics.dk/~amoeller/XML/querying/index.html (een korte Xquery-tutorial)
www-106.ibm.com/developerworks/xml/library/x-xquery.html (artikel *An introduction to Xquery*)

Leveranciers

www.softwareag.com/developer/quip
www.x-hive.com/xquery
<http://131.107.228.20/xquerydemo> (Microsoft)
www.ibm.com/developerworks/
www.xmlglobal.com/prod/db/
<http://demo.openlinksw.com:8891/xquery/demo.vsp> (Virtuoso Server)

Open source

<http://sourceforge.net/projects/xquen/>
www.fatdog.com (XQEngine)
www.gnu.org/software/kawa/xquery (Kawa-Query)

Onderzoeksprojecten

<http://db.bell-labs.com/galax/>
<http://xml.darmstadt.gmd.de/xquerydemo/> van (IPSI-XQ).

(codenaam Yukon, voorzien voor 2003) native XML-ondersteuning zal aanbieden, met inbegrip van het gebruik van XQuery als query-taal.

Daarnaast is IBM druk bezig met de ontwikkeling van een native XML-database (codenaam Xperanto), die zal dienstdoen als een onderdeel van DB2. Een eerste pre-bètaversie van Xperanto zal in de eerste helft van dit jaar openbaar gemaakt worden via IBM's developer-website.

Voor XML-databases dreigt hetzelfde lot als objectgeoriënteerde databases

Verder is er nog een aantal andere commerciële XML-databases, zoals GoXML van XML Global of Virtuoso Server van Openlink, die XQuery beginnen te ondersteunen, weliswaar elk in een nét iets verschillende, iets van de voorlopige standaard afwijkende vorm.

Speelt u liever met open source-software, dan hebt u hier ook al de keuze uit

een aantal Xquery-implementaties. Zo vindt u bij SourceForge XQuench een in Java geschreven Xquery-processor, die zich echter nog maar in een eerste prille alfarelease bevindt. Gelijksortige programma's zijn XQEngine en Kawa-Query.

Ten slotte loopt een aantal onderzoeksprojecten waarin de XQuery taal verder verfijnd wordt, zoals Galax van Bell Labs en IPSI-XQ van GMD-IPSI.

CONCLUSIES

XQuery heeft duidelijk nog een lange weg te gaan alvorens het serieus aanspraak kan maken op de titel van volwaardige XML-datamanipulatietaal. Ofschoon hiërarchische gegevensstructuren, zoals XML die kent, op het eerste zicht eenvoudiger lijken dan relationele, is XQuery snel aan het uitgroeien tot een taal die veel complexer is dan SQL en toch veel minder mogelijkheden biedt. Bovendien is de taal zo ingewikkeld aan het worden, dat te vrezend valt dat alleen de grote databaseleveranciers nog in staat zullen zijn hiervan

een efficiënte implementatie te bouwen. Gelukkig is er inmiddels toch een eerste werkversie van Xquery, die binnenkort wel in de meeste native XML-databaseproducten ondersteund zal worden. Een hele verbetering vergeleken met de wirwar aan vendor-specifieke query-talen en API's die deze producten momenteel gebruiken.

Zolang er geen echte insert/update/delete-functionaliteiten in XQuery worden voorzien, lopen deze XML-databases echter het risico hetzelfde lot beschoren te zijn als objectgeoriënteerde databases. Technisch superieur voor een aantal gespecialiseerde toepassingen, maar qua bruikbaarheid en betrouwbaarheid duidelijk inferieur aan traditionele rdbms'en, waaraan nu in snel tempo goede XML-ondersteuning wordt toegevoegd. ●

Hans C. Arents (hca@itworks.be; www.arents.net) is onafhankelijk XML-technologieadviseur bij I.T. Works, een organisator van product- en leverancieronafhankelijke IT-seminars in België.

Vervolg van pagina 15

Daarbij wordt de inhoud van de oude database uitgelezen en terug ingelezen in de nieuwe database.

Gezien het verhoogde alfa-gehalte van auteur dezes is het wijzigen van een XML-schema daarentegen een normale zaak. De kans is namelijk groot dat de bestaan-

Dankzij de volledigheid van de schemata zal niet zo gauw een wildgroei aan DDL-varianten ontstaan

de dataset best kan blijven voldoen aan een nieuwe versie van het schema. Een XND is dus te prefereren boven een rdbms als men verwacht dat het datamodel aan wijziging zal onderhevig zijn.

Ook voor semi-gestructureerde of documentgeoriënteerde datasets waarbij men de hiërarchie binnen de documenten wenst te ontsluiten, geeft een XND zijn relationele broertje het nakijken. Bij toe-

passingen als document of website content management verdient het zonder meer de aanbeveling de duur vergaarde content niet te laten verwateren als een CLOB in een relationeel model. Al helemaal te gek zou een poging zijn metastructuren, namespaces en hypertextstructuren te vatten in het keurslijf van het E/R-model. Het bestaan van een markt voor leveranciers van complexe middleware, die de mapping mogelijk maakt tussen objecten en relationele databases, is mede het bewijs voor de noodzaak van een persistentieformaat voor dit soort objectstructuren in een sequentieel-hiërarchisch kleedje. Daarvoor zou XND ontegenzeggelijk beter geschikt zijn.

Ten slotte: XND's hebben de juiste interfaces voor het vlot aan- en afkoppelen binnen losjes gekoppelde infrastructuurkettingen. Een http-interface, eenvoudige (soms simplistische) query-faciliteiten aan de hand van XPath en een gedoogbeleid voor het ontbreken of frequent wijzigen van het datamodel -het is immers niet het business-datamodel, maar het generieke XML-model dat deze systemen beheren-

maken een XND uitermate geschikt als frontend-technologie voor de aggregatie en eenduidige representatie van de resultaten van complexe business-processen op de backend. ●

Noot:

1. Zie www.firstsql.com/dbdebunk/xmldev1.htm
2. Microsoft beweert zelfs dat de volgende versie van SQL Server "in niets" te onderscheiden zal zijn van een XML-native database (XND).
3. In het thema-artikel *XQuery zet zijn eerste schuchtere stappen*, elders in dit nummer, schrijft Hans Arents wat er in *fundamentele* zin nog schort aan XND's om ze een volwaardig dbms te kunnen noemen.
4. Zie www.sees.bangor.ac.uk/~rich/research.html
5. Zie www.w3.org/TR/xml-infoset/

Steven Noels (stevonn@outerthought.org) is mede-oprichter van Outerthought, een klein en spitstechnologisch kenniscentrum voor Java en XML. Als productmanager bij E-corporation introduceerde hij verschillende open source-toepassingen. Daarvoor was hij als key technical SGML/XML-consultant en tool-specialist.