

Verschillende interpretaties voor de tuples van een relatie

Over Polariteit, Modaliteit en de Open World Interpretatie

Maurice Gittens

Codd stelde in 1979 een relationeel model voor waarin plaats is gemaakt voor de zogenaamde Open World Interpretation van databaseproposities. In dit artikel worden enkele concepten geïntroduceerd die aansluiten bij de Open World interpretatie van gegevens in relationele databases. De Open World interpretatie verbetert de getrouwheid waarmee informatiemodellen een discussiedomein reflecteren.

Een relationele database is te zien als een verzameling gestructureerde proposities over een bepaald discussiedomein, die in essentie dient als communicatiemiddel tussen (bedrijfs)processen die binnen dat discussiedomein plaatsvinden. In het algemeen geldt dat de kwaliteit van de informatievoorziening verbeterd wordt wanneer de proposities met meer getrouwheid een reflectie zijn van het discussiedomein. In dit artikel worden de concepten Polariteit, Modaliteit en Volledigheid inleidend onder de aandacht gebracht.

Deze concepten worden relevant als we kiezen voor een Open World interpretatie voor databasegegevens. Voordat we ingaan op deze concepten plaatsen we ze in de context van het relationele model volgens Codd en datzelfde model volgens Date en Darwen.

Closed of Open World Interpretation

In het artikel "Extending the relational database model to capture more meaning" [3] stelt Codd:

One may choose to interpret the absence of an admissible tuple from a base relation as a statement that the truth value of the corresponding atomic formula is (1) unknown; (2) false. If (1) is adopted, we have the Open World interpretation. If (2) is adopted, we have the Closed World interpretation (see [28]).

De Closed World interpretatie is een vrije vertaling van de stellingname: "Alles wat we niet weten is niet waar". Blijkens de uiteenzetting op pagina 29 van The Third Manifesto [4] hangen Date en Darwen de Closed World Interpretation aan. Ik citeer:

Furthermore, we subscribe, noncontroversially, to the Closed World Assumption [121] which says that if a given tuple plausibly

could appear in the relvar at some time but in fact does not, then the corresponding proposition is understood by convention to be one that evaluates to FALSE at the time in question.

Elders in het bovengenoemde artikel stelt Codd:

In this treatment, if an operator generates one or more nulls, these nulls are always of the type "value at present unknown", which is consistent with the Open World interpretation (see Section 3). If we were dealing with relations having a Closed World interpretation, the "property inapplicable" type would be appropriate.

De modaliteit van specifieke relaties zal in het algemeen in de catalogue gedefinieerd worden

Dit citaat geeft inzicht in het streven van Codd naar consistentie ook in relatie tot de Open World interpretatie van databaserelaties. Uit symmetrie-overwegingen leidt de onderkenning van "unknown" met betrekking tot tuples al snel tot de onderkenning van "unknown" in relatie tot attribuutwaarden. Is dan de opvatting van Codd met betrekking tot NULL niet te herleiden tot zijn onderkenning van de Open World interpretatie? Ik denk van wel.

Nu we een inleidend beeld hebben van de achtergronden gaan we in op verschillende interpretaties van relaties.

Polariteit van tuples in relaties

De tabel in afbeelding 1 is een weergave van een relatie die een verzameling proposities vastlegt over een aantal individuen en hun werkgevers. Deze relatie¹ legt volgens de klassieke interpretatie bijvoorbeeld vast:

Het is waar dat: Marta een Persoon is die als Werkgever IBM heeft.

In de terminologie van functionele linguïsten [1] gebruikt deze interpretatie de *positieve* polariteit. We nemen namelijk aan dat

Persoon-x-Werkgever

Naam	Werkgever
Marta	IBM
Piet	Google
James	Microsoft

Afbeelding 1: Tabel.

Iedere tuple in een relatie een propositie weergeeft die waar is. Een interpretatie met negatieve polariteit zou zijn:

Het is niet waar dat: Marta een *Persoon* is die als *Werkgever* IBM heeft.

Als we onderkennen dat:

- we vaak niet weten of een propositie waar is of niet;
 - het er vaak niet toe doet of een propositie waar is of niet;
- dan kunnen we ook de neutrale polariteit onderscheiden.

De bewering wordt dan:

Er is bewerd dat: Marta een *Persoon* is die als *Werkgever* IBM heeft.

De polariteit van de tuples van specifieke relaties zal in het algemeen in de catalogue gedefinieerd worden.

Modaliteit van tuples relaties

In de formele logica en ook in de functionele linguïstiek wordt het begrip modaliteit gehanteerd. Modale interpretaties van de voorbeeldpropositie zijn bijvoorbeeld:

Het is noodzakelijk waar dat: Marta een *Persoon* is die als *Werkgever* IBM heeft.

Het is mogelijk waar dat: Marta een *Persoon* is die als *Werkgever* IBM heeft.

De modaliteit van specifieke relaties zal in het algemeen in de catalogue gedefinieerd worden.

Polariteit en Modaliteit van attributen van relaties

Als in onze voorbeeldrelatie het attribuut werkgever de negatieve polariteit zou worden toegekend dan zou de interpretatie van een tuple in deze relatie kunnen luiden:

Het is waar dat: Marta een *Persoon* is die *niet* IBM als *Werkgever* heeft.

De modaliteit en polariteit van specifieke attributen zullen in het algemeen in de catalogue gedefinieerd worden.

Polariteit en Modaliteit van individuele tuples en attribuutwaarden

Modaliteit en polariteit kunnen ook door middel van relationele beginselen voor individuele tuples en specifieke attribuutwaarden worden gedefinieerd. Op basis van beginselen die Codd [3] heeft geïntroduceerd kan polariteit en modaliteit op flexibele wijze ook op het niveau van individuele tuples en specifieke attribuutwaarden worden gedefinieerd. Dit kan door de inzet van specifieke orthogonalen (zie DB/M 6, jaargang 2005 en ook DB/M 7, 2006 – Hogere orde concepten en hun visualisatie). Hierop gaan we in dit artikel niet verder in.

Logische connectoren voor proposities: de conjunctieve interpretatie

In het algemeen is de interpretatie van een relatie conjunctief. Dit wil voor ons voorbeeld zeggen dat de verschillende proposities in onze relatie met een logische 'en' operator worden verbonden. Dus de interpretatie wordt:

Het is waar dat: Marta een *Persoon* is die als *Werkgever* IBM heeft *en*

Het is waar dat: Piet een *Persoon* is die als *Werkgever* Google heeft *en*

Het is waar dat: James een *Persoon* is die als *Werkgever* Microsoft heeft.

Het is ook mogelijk om de verschillende proposities met de logische 'of' operator te verbinden. De interpretatie van de relatie is dan disjunctief:

Het is waar dat: Marta een *Persoon* is die als *Werkgever* IBM heeft *of*

Het is waar dat: Piet een *Persoon* is die als *Werkgever* Google heeft *of*

Het is waar dat: James een *Persoon* is die als *Werkgever* Microsoft heeft.

Deze interpretatie impliceert dat minimaal een en maximaal drie van de bovenstaande proposities waar zullen zijn in het discussiedomein. Terwijl bij de conjunctieve interpretatie alle drie de

Seizoen-x-Korting

Seizoen	Korting (%)
Winter	8
Lente	5
Zomer	0
Herfst	5

Afbeelding 2: Tabel.

proposities waar zullen zijn. Het is natuurlijk ook mogelijk om een tabel te interpreteren met de exclusieve-disjunctie operator; ook bekend als de 'exclusive or'. Een interpretatie van onze voorbeeldrelatie op basis van exclusieve-disjunctie kan luiden:

Het is exclusief waar dat: Marta een Persoon is die als Werkgever IBM heeft of

Het is exclusief waar dat: Piet een Persoon is die als Werkgever Google heeft of

Het is exclusief waar dat: James een Persoon is die als Werkgever Microsoft heeft.

Deze interpretatie geeft aan dat maximaal een van de bovenstaande proposities op een gegeven moment waar mag zijn. Beschouw als voorbeeld eens de relatie die een korting koppelt aan een seizoen, zoals in de tabel in afbeelding 2.

De exclusieve disjunctie zouden we kunnen gebruiken om te benadrukken dat op een gegeven moment een (1) kortingspercentage mag gelden. Door in de metadata van deze relatie aan te geven dat hij op basis van de exclusieve disjunctie operator geïnterpreteerd dient te worden, wordt expliciet gemaakt dat er op een gegeven moment precies een van de bovenstaande proposities waar zal zijn.

De volledigheid van een relatie

Een relatie kunnen we zien als een verzameling proposities met dezelfde structuur. Bij een Closed World interpretatie geldt de aanname dat de verzameling volledig is, terwijl bij een Open World interpretatie niet bekend is of de verzameling wel of niet volledig is. De eenduidigheid van een gegevensmodel wordt verbeterd als de metadata van een relatie aangeven of de relatie volgens het model wel of niet volledig is.

Streven naar de eenduidige interpretatie van gegevens

Het is belangrijk om bij het ontwerpen van informatiemodellen zoveel mogelijk eenduidig te zijn met betrekking tot de bedoelde interpretatie van de gegevens. Voor iedere relatie zijn de antwoorden op de volgende vragen relevant (niet limitatief):

- is de bedoelde interpretatie conjunctief, disjunctief of exclusief disjunctief;
- is de relatie volledig;
- wat is de polariteit van de proposities;
- wat is de modaliteit van de proposities?

Veelal wordt informatie over de correcte interpretatie van databasegegevens niet in de database ondergebracht maar in applicaties en/of geassocieerde documentatie. Deze praktijk leidt tot een vermindering van de kwaliteit van de informatievoorziening. Als u weet waarom u integriteitsregels en bedrijfsregels zoveel mogelijk in de database dient onder te brengen, dan weet u ook waarom gegevens over de correcte interpretatie van gegevens idealiter in de database thuis horen. Relationale data-

bases onderscheiden zich niet alleen door de inzet van de relationele algebra maar ook door de zelfbeschrijvendheid. De mate waarin de zelfbeschrijvendheid dekkend is wordt verbeterd door interpretatieregels voor databasegegevens in de database zelf onder te brengen.

De Open World interpretatie voor database gegevens accepteert en erkent de realiteit dat er veel bestaat dat niet bekend is. Codd heeft in zijn relationele model de erkenning van deze realiteit gefaciliteerd. De Open World interpretatie van gegevens is naar mijn mening in het algemeen te verkiezen boven de Closed World interpretatie en dan wel omdat de Open World interpretatie kwalitatief betere informatiemodellen oplevert.

Hoe zit het dan met NULL's en Three Valued Logic?

Het is bekend dat NULL's zoals deze doorgaans in SQL-dialecten zijn gedefinieerd aan veel kritiek worden blootgesteld. Over de vraag of deze kritiek naar mijn mening wel of niet terecht is wil ik het volgende zeggen:

- is het onderwerp NULL's in een discussie over informatiemodelleren niet een digressie;
- is het gebruik van NULL's in een implementatiemodel relevant op logisch en conceptueel niveau;
- is het feit dat NULL's in implementaties maar al te vaak misbruikt worden een reden om deze in conceptuele zin te verwerpen?

Een relatie is te zien als een verzameling proposities met dezelfde structuur

De antwoorden op deze en soortgelijke vragen leiden mij tot de conclusie dat NULL's, in de context van SQL, hun plaats hebben in de reeks van hulpmiddelen die ter beschikking staan van de database professional. Ook aan de toepassing van Three Valued Logic (3VL) kleeft, mijn inziens, niets dat inherent verkeerd is. De vuistregel die ik zelf gebruik luidt:

Model the concept as accurately as practically possible.

Een conceptueel zuiver model leidt doorgaans tot implementaties die de conceptuele integriteit van het model reflecteren. Wat er omwille van bijvoorbeeld de performance en de implementatie verder in het implementatiemodel wordt gedaan dient goed aan te sluiten bij wat opportuun is. Deze vuistregel impliceert doorgaans dat het onbekende en dus ook 3VL alleen daar gebruikt dient te worden waar het in praktische zin niet te vermijden is.

Conclusie

Dit artikel heeft:

- de Open World en de Closed World interpretatie onder de aandacht gebracht;
- onderbouwd dat het relationele model volgens Codd de Open World interpretatie faciliteert;
- onderbouwd dat het relationele model volgens Date en Darwen uitgaat van de Closed World interpretatie;
- aan de hand van eenvoudige voorbeelden de concepten Polariteit en Modaliteit onder de aandacht gebracht als concepten die passen bij een Open World interpretatie;
- de conjunctieve, disjunctieve en exclusief disjunctieve interpretatie van relaties onder de aandacht gebracht als concepten die passen bij de Open World interpretatie;
- gesteld dat de Open World interpretatie de voorkeur geniet, omdat informatiemodellen op basis van deze interpretatie in het algemeen met meer getrouwheid het discussiedomein kunnen modelleren;
- gesteld dat een relationele database te zien is als een verzameling gestructureerde proposities over een bepaald discussie-

domein, die in essentie dient als communicatiemiddel tussen (bedrijfs)processen en deelnemers aan deze processen;

- gesteld dat het streven naar eenduidige interpretaties van database gegevens wenselijk is.

Noten

1. *De relaties Persoon en Werkgever met corresponderende foreign-key relaties laten we voor het gemak achterwege.*

Literatuur

1. M.A.K. Halliday; 2004; *An Introduction To Functional Grammar*; Arnold.
2. Jan van Eijck, Elias Thijsse; 1989; *Logica voor alfa's en informatici*; Academic Service.
3. E.F. Codd; 1979; *Extending the relational database model to capture more meaning*; IBM.
4. Date, Darwen, 2007; *Databases, Types and the Relational Model*, Addison-Wesley Longman.
5. E.F. Codd; 1990; *The Relational Model for Database Management*; Addison-Wesley.

Maurice Gittens is zelfstandig IT-consultant.



BI-ware

De harde en de zachte kant van Business Intelligence

BI-initiatieven mislukken nog veel vaker dan andere projecten.

De BI-initiatieven moeten van de harde en de zachte kant komen. En als de harde kant van BI al praat met de zachte, spreken ze niet elkaars taal.

Het boek BI-ware is een boek voor ICT'ers en voor managers en vertelt in gewoon Nederlands wat er allemaal fout kan gaan en wat daaraan kan worden gedaan.

BI-ware bevat een bundeling van artikelen van Karien Verhagen en is een nieuwe uitgave in de reeks van DB/M Essays. De artikelen zijn gepubliceerd in de periode 2002 – 2006.

Wilt u weten hoe u Business Intelligence kunt laten slagen?

Dan kunt u niet zonder deze uitgave!

Ga snel naar www.array.nl en bestel BI-ware!

Deze uitgave is mogelijk gemaakt door: **Getronics** **PinkRocade**

DB/M

Array PUBLICATIONS