

Ontrafeling van de Babylonische spraakverwarring

Het Metadata Frame in Actie (deel 2)

Peter Alons

In Database Magazine 6 is beschreven hoe bij het Erasmus Medisch Centrum in Rotterdam gebruik wordt gemaakt van de Atos Origin Metadata Frame methode. Dit is het tweede deel.

Voor de feiten uit afbeelding 2 waren de medische experts het erover eens dat het allemaal voorbeelden zijn van een feittype dat bestaat uit een real-time meting en het resultaat daarvan. De real-time meting is de meting van een real-time variabele, geïdentificeerd door zijn naam, op een gegeven moment in het bestaan van een patiënt. Het resultaat bestaat uit een waarde met een geschikte eenheid. Afbeelding 5 toont de ontledingsboom of 'expression tree' voor dit feittype in CaseTalk [5], de case-tool die we gebruiken voor de analyse van de feiten. Deze ontleding bepaalt grotendeels de structuur van het op te leveren model. Meer informatie over dit proces is te vinden in [4]. Zodra alle relevante feiten in de sheets met concrete voorbeelden verwoord zijn in zinnen goedgekeurd door de medische experts, en op dezelfde manier zijn geanalyseerd als in afbeelding 5, kan CaseTalk uit al deze feiten een volledig informatiemodel afleiden met behulp van een exact algoritme. Dit algoritme wordt gedicteerd door 'predicatenlogica', de wiskundige theorie achter FCO-IM. Het produceert een relationeel model dat aan alle normalisatieregels van Codd voldoet met het minimum aantal tabellen dat nodig is om alle informatie te bevatten die correspondeert met de relevante feiten: niets meer en niets minder. Het

feittype in afbeelding 5 verschijnt in dit model als onderdeel van de tabel getoond in afbeelding 6. In afbeelding 6 komt het laatste veld in de tabel Real_Time_meting voort uit een feittype waarvan de eerste zin blijkbaar was: *Is het resultaat voor Hartfrequentie bij Patiënt P1234567 op 21-11-2005 om 9:00:00 uur gevalideerd?* 1. Deze zin verwoordt een feit getoond op een niet zichtbaar deel van de voorbeeldsheet in afbeelding 2, waar was aangegeven dat de dienstdoende verpleegkundige staf deze gegevens gevalideerd had.

Zoals blijkt uit de foreign-key toewijzingen in afbeelding 6, produceert het algoritme ook de tabellen Patiënt, Dag, en Tijdstip die voorkomen in de expressieboom in afbeelding 5. Deze tabellen hebben als primary-key Patiëntnummer, Datum, en Tijd, en bevatten vele andere velden voortvloeiend uit relevante feiten in alle verzamelde voorbeelddocumenten. Niet verrassend maar beslist plezierig is het feit dat de tabellen Patiënt, Dag, en Tijdstip ook voorkwamen als kandidaat-dimensies in de spreadsheet van fase 1 van onze methode. Daar werd de tabel Tijdstip simpelweg aangeduid met Tijd. In dezelfde geest leiden analyses als in afbeelding 5 van de feittypen die horen bij de figuren 3 en 4 tot de tabellen 'Laboratorium_meting' en 'Beademings-uurfragment'.

Perfect model

Als CaseTalk zijn algoritme toepast wordt een 'logisch datamodel' in optimale normaalvorm geproduceerd voor de Corporate Fact Base, maar dit zit nog steeds in de repository van CaseTalk.



Afbeelding 5: 'Expression tree' in CaseTalk voor het feittype "Real-time meting met resultaat".

RT_variabelenaam (RT_variabelenaam) NN PK	Patiëntnummer (Patiëntnummer) NN PK	Datum (Datum) NN PK	Tijd (Tijd) NN PK	Waarde (Waarde) NN	Eenheid (Eenheid) NN	Gevalideerd (Boolean) NN
<i>Hartfrequentie</i>	<i>P1234567</i>	<i>21-11-2005</i>	<i>9:00:00</i>	<i>112</i>	<i>slagen / min.</i>	<i>1</i>
<i>Temperatuur</i>	<i>P1234567</i>	<i>21-11-2005</i>	<i>7:00:00</i>	<i>38.2</i>	<i>°C</i>	<i>1</i>

FEITTYPE EXPRESSIONS:
 F1: "Er is bij Patiënt <Patiëntnummer> op <Datum> om <Tijd> uur voor <RT_variabelenaam> een waarde van <Waarde> <Eenheid> vastgelegd."
 F2: "Is het resultaat voor <RT_variabelenaam> bij Patiënt <Patiëntnummer> op <Datum> om <Tijd> uur gevalideerd? <Gevalideerd>."

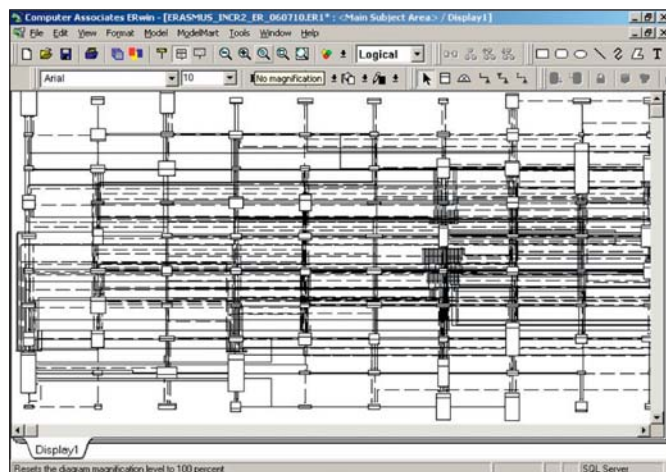
FOREIGN KEYS:
 Real_Time_meting (RT_variabelenaam) → RT_variabele (RT_variabelenaam)
 Real_Time_meting (Patiëntnummer) → Patiënt (Patiëntnummer)
 Real_Time_meting (Datum) → Dag (Datum)
 Real_Time_meting (Tijd) → Tijdstip (Tijd)

Afbeelding 6: De tabel Real_Time_meting voortvloeiend uit het model creërend algoritme van CaseTalk.

Een ander hulpmiddel in de volledige toolset van het Metadata Frame, de ERM-Bridge, kan het model in zijn geheel vanuit CaseTalk naar diverse ER/UML-omgevingen transporteren. Dit hulpmiddel is oorspronkelijk ontwikkeld op de Hogeschool van Arnhem en Nijmegen, die nog steeds de bakermat van FCO-IM is en onze partner bij het uitdenken van vele van de ideeën die in ons Metadata Frame worden toegepast.

Na het transporteren van het model, is dit beschikbaar in de gekozen ER-tool zonder verlies van enige belangrijke informatie over het model. In het bijzonder is alle semantiek die aan het model ten grondslag ligt, overgebracht in de ER-repository in de vorm van commentaren bij de attributen. De precieze betekenis van elk veld in het model kan daarom nog steeds begrepen worden in termen van de oorspronkelijke zinnen.

Het volledige model in de ER-tool, dat door de algoritmen is gegenereerd, is zichtbaar in afbeelding 7. En inderdaad, het is een heel groot model. Dit is precies waarom het niet met de hand moet worden gemaakt in IT-gedreven top-down benadering, waarbij de experts niet serieus betrokken zijn. Zelfs in het onwaarschijnlijke geval dat een dergelijke benadering tot een volledig correct model van de communicatie leidt, zijn mensen niet in staat het model volledig correct in te voeren. Een



Afbeelding 7: Het volledige ER-model voor het Erasmus MC zoals het door de algoritmen is gegenereerd.

dergelijk model kan beter gegenereerd worden door een strak algoritme vanuit een conceptueel model dat volledig gevalideerd is door experts, zoals in onze aanpak het geval is. Het algoritme is niet in staat om fouten te maken. Daarom weten we domweg dat het resulterende ER-model goed is. Of zelfs beter: dat het *perfect* is.

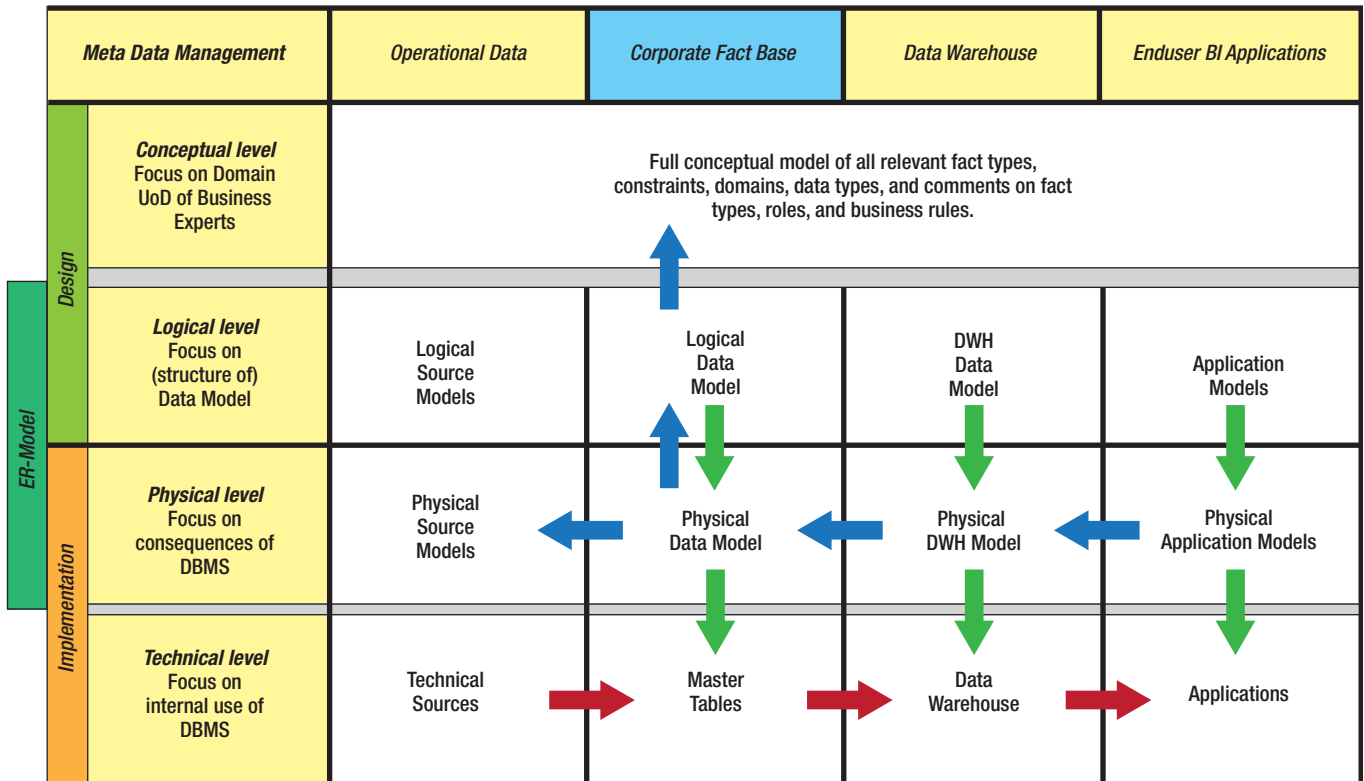
Ontrafeling

In principe kan onze modelleringsaanpak in afbeelding 1 (zie DB/M 6) ook worden gezien als een 'Plan-Do-Check-Act' cyclus. Zodra deze cyclus is afgerond, kunnen we de modellen creëren die in de datawarehouse-omgeving geïmplementeerd worden. De feitelijke validatie van deze modellen is dan al achter de rug. Dit betekent, dat onze Metadata Frame methode voldoet aan de volgende eisen:

- Doelgerichtheid.** We modelleren niet de werkelijkheid, maar de *communicatie* van de business experts over hun werkelijkheid. Dit resulteert in feitexpressies in de taal van de business experts zelf. We modelleren alle conceptuele aspecten van de communicatie *zonder vooroordelen of beperkingen*;
- Doeltreffendheid.** De business experts zijn in staat te valideren dat hun communicatie volledig en correct gemodelleerd is, zonder zelf naar enig gegenereerd datamodel te kijken *en zonder hun kostbare tijd te verspillen aan voor hen irrelevante en onbegrijpelijke technologische aspecten*;
- Doelmatigheid.** We slaan alle modelinformatie op in één generieke repository voor alle niveaus, en gebruiken repository-transformaties om alle metadata via algoritmen te transporteren naar iedere plek waar deze nodig is voor verdere naslag.

Conclusie

Als IT'ers hebben wij in dit proces de rol van leverancier, en met onze methode tonen we volledig respect voor de business experts als onze klanten. Doordat we in het ontwerpproces volledige validatie verkrijgen, creëren we volledige klanttevredenheid. Dit leidt tot minimale verspilling van tijd en hulpmiddelen in de implementatiefase. Dat blijkt uit de volgende uitspraken gedaan door onze collega's, die het datawarehouse bij het Erasmus MC op basis van ons ontwerp gebouwd hebben:



Afbeelding 8: Het Metadata Frame en zijn functioneel gebruik in de implementatieprocessen.

- "Het model is volledig begrijpelijk voor alle afnemers, eindgebruikers en ontwikkelaars gelijk: de zinnen en voorbeelden maken het eenvoudig te communiceren over de te leveren functionaliteit."
- "De ontwikkelaars worden door de voorbeelden sterk ondersteund bij de bouw: de voorbeelden prikkelen hun voorstellingsvermogen op de juiste manier."
- "De gehele toolset van het Metadata Frame werkt als communicatiemiddel tussen de medische experts en de IT-staf."
- "De metadata zijn moeiteloos beschikbaar te maken bij de attributen zowel in de ETL als in de front-end applicaties."
- "De gemodelleerde stermodellen bevatten *alle* rapportagebehoefte: het bouwen van de Business Objects universes is kinderspel."

Voor wie het Metadata Frame kent, is dit geen verrassing: alle rapportagebehoefte dienden als invoer voor een proces dat volledige validatie garandeert. Verder zijn de functionele stappen in het implementatieproces (zie de pijlen in afbeelding 8) maximaal geautomatiseerd, waarmee de kans op fouten is geminimaliseerd. De blauwe pijlen geven de processen weer, die door onze metadata tool MetaMorfose worden uitgevoerd. Deze tool legt automatisch alle informatie over de modellen vast, die voor de diverse zuilen zijn ontworpen, en koppelt deze modellen groten-deels geautomatiseerd aan elkaar. Dit ondersteunt krachtig de rode pijlen, die de ETL-processen weergeven, en evenzo de geautomatiseerde levering van relevante metadata aan eindgebruikers in hun front-end applicaties. Dit geldt in het bijzonder

voor alle semantiek die aan de velden in deze applicaties ten grondslag ligt. Zoals we lieten zien in [6] hebben de zinnen die deze velden genereerden veel meer betekenis voor een eindgebruiker dan standaarddefinities van deze velden. Definities neigen de aandacht te trekken naar de *geldigheid* van de inhoud van velden, terwijl de concrete semantiek de aandacht trekt naar de *juistheid* van deze inhoud. En gezien vanuit informatie-kwaliteit is het laatste veel belangrijker dan het eerste. Toen de eerste resultaten van de applicatieontwikkeling beschikbaar kwamen, verzekerden de betrokken medische experts ons al spoedig dat zij met deze applicaties alle wensen met betrekking tot hun informatiebehoefte konden vervullen en dat zij er erg verguld mee waren. Dit werd nog verder opgeluisterd door de gelukkige uitkomst dat het Erasmus MC en Atos Origin met dit project de Computable Award 2007 gewonnen hebben voor het beste IT-project in de Zorg.

Literatuur

1. *Volledig Communicatiegeoriënteerde Informatiemodellering*, G. Bakema, J.P. Zwart, H. van der Lek, Kluwer BedrijfsInformatie, 2002.
2. *CaseTalk*, Bommeljé, Crompvoets, en Partners, www.casetalk.com.
3. *Database Magazine*, Jaargang 17, nr. 2 (april 2006), P.W.F. Alons en R.G. Arntz.

Peter Alons

Dr. P.W.F. Alons (peter.alons@atosorigin.com) is senior consultant bij Atos Origin/BI-CRM en al ruim vijftien jaar betrokken geweest bij een groot aantal BI- en Datawarehouse-projecten bij diverse bedrijven in diverse branches.