

DATAMODELLERING BEGRIPPENBOOM

Inleiding

In dit whitepaper wordt de datamodelleervorm begrippenboom inclusief de begrippenlijst beschreven. Deze modelleervorm staat in verhouding tot een aantal andere modelleervormen. Wil je een beeld krijgen van welke modelleervormen er zijn bekijk dan het whitepaper wat een introductie geeft tot datamodelleervormen en deze serie van whitepapers zie Inleiding datamodellering

Doel

De begrippenboom heeft tot doel om op semantisch niveau de data- of bedrijfsconcepten binnen een domein in kaart te brengen. Dit in kaart brengen bestaat allereerst uit het opstellen van een lijst van begrippen met de daarbij behorende definities of beschrijvingen. Met name in omvangrijke domeinen zoals grote organisaties of clusters van verschillende organisaties is het opstellen van een begrippenlijst (in het Engels veelal Glossary genoemd) een complexe activiteit.

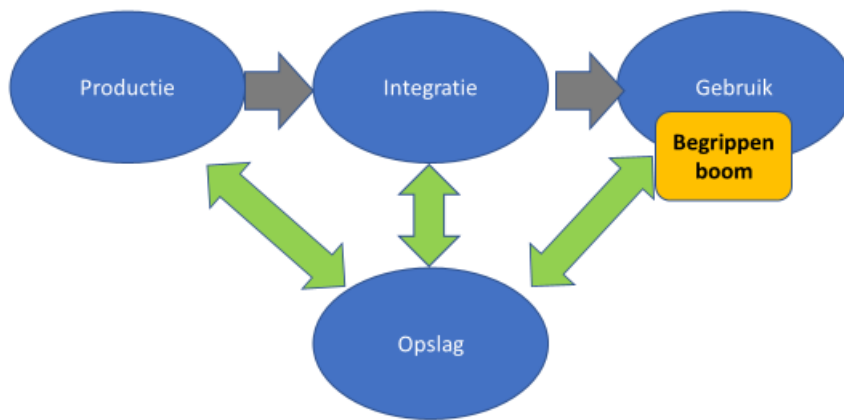
Is een begrippenlijst opgesteld dan kan vervolgens een begrippenboom opgesteld worden waarbij verschillende hiërarchieën en verbanden tussen de begrippen gemodelleerd kunnen worden.

Het doel van een begrippenboom is om op semantisch vlak overeenkomst te bereiken bij de betrokken stakeholders wat de betekenis of de definitie van een begrip is. Vandaar dat de combinatie van begrip en beschrijving van het begrip een randvoorwaarde is voor een begrippenboom en een begrippenlijst.

Raamwerken

DATALEVENSLLOOP

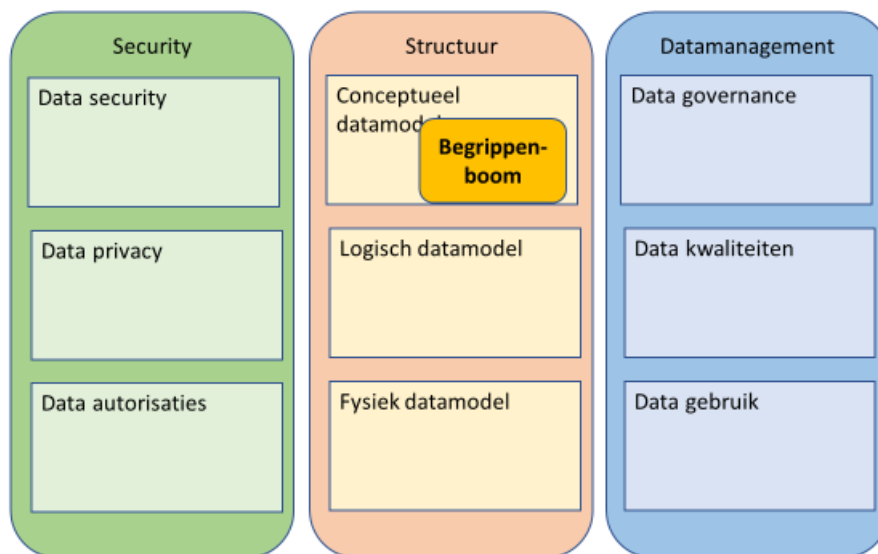
In de datalevensloop wordt de begrippenboom met name gebruikt binnen het datagebruik perspectief. Hier dient overeenstemming te zijn over de begrippen die bij het gebruik (en met name bij het nemen van besluiten op basis van de gebruikte datasets) relevant zijn. Is er geen overeenstemming over de begrippen dan zal op zeker moment in het meest positieve geval discussie ontstaan rond de gegevens die gebruikt worden om data om te zetten naar kennis. In het meest negatieve geval worden de begrippen verkeerd geïnterpreteerd en ontstaat de kans dat er verkeerde beslissingen genomen worden. Stel bijvoorbeeld dat in een ziekenhuis het begrip patiënt en de definitie niet duidelijk is, de gevolgen kunnen dan verstrekkend zijn.



In de afbeelding is te zien dat de begrippenboom slechts in een klein deel, met name datagebruik, een rol speelt. Afgeleid kan het een rol spelen in productie omdat dit regelmatig beïnvloed wordt door het datagebruik omdat het gebruik eisen stelt aan de product, echter in deze situatie wordt dit niet nader uitgewerkt.

DATARAAMWERK

Ook in het dataraamwerk wordt de begrippenboom beperkt ingezet, alleen in de conceptuele modellering. Veelal wordt als eerste begonnen met het opstellen van een begrippenboom als conceptueel model, bijvoorbeeld bij een top down aanpak.



In een aantal gevallen wordt de begrippenboom breder ingezet, bijvoorbeeld als glossary voor project documentatie. In dat geval zal de begrippenlijst meer begrippen bevatten dan vanuit datamodelleringsperspectief noodzakelijk is. Echter ook in dat geval blijven de begrippen op conceptueel niveau uitgewerkt.

Stakeholders

Bij het opstellen van een begrippenboom is vertegenwoordiging vanuit de business noodzakelijk. Data architecten kunnen ondersteunen bij het opstellen van een begrippenboom maar zullen nooit in staat zijn om zelf een adequate definitie op te stellen. Onderstaande opsomming geeft de belangrijkste stakeholders:

- **Domeinexperts**, de belangrijkste stakeholder, zij dienen aan te geven welke begrippen relevant zijn en met elkaar in discussie gaan om tot een geaccepteerde definitie van een begrip en de onderlinge verhouding van de begrippen te komen
- **Gebruikers**, in kleinere domeinen, bijvoorbeeld waarbij rapportages en business intelligence ingezet wordt kunnen gebruikers van bijvoorbeeld rapportages ingezet worden voor het verfijnen en uitbouwen van een begrippenboom
- **Data analisten en scientists**, een relatief nieuwe groep stakeholders, op basis van datasets gaan deze stakeholders analyses doen met datasets en de gebruikte begrippen. Zij bouwen hierdoor diepgaande kennis op van de begrippenboom en de achterliggende objectmodellen.
- **Functioneel (Applicatie) Beheerders**, een groep stakeholders die diepgaande kennis heeft van (standaard) informatiesystemen en het onderliggende datamodel inclusief de door de leverancier gebruikte begrippen.
- **Data of Informatiearchitect**, zal veelal beperkt kennis hebben van de inhoudelijke definities van de in het domein gebruikte begrippen. Echter deze stakeholders hebben veelal wel diepgaande

kennis rond data patronen, diagnostische technieken of het opstellen van modellen rond begrippenbomen.

Concepten

Binnen de begrippenlijst en de begrippenboom staan twee zaken centraal, namelijk:

- **Begrip**, de naam zoals dit binnen de domeinpopulatie gebruikt wordt
- **Definitie**, omschrijving van het begrip waarover overeenstemming is binnen de populatie. Met name die overeenstemming kan een uitdaging zijn. In dat geval zie je nog wel eens een definitie die zo algemeen is dat het altijd geldend is. Dat zal in een later stadium discussie of problemen gaan geven bij de implementatie van het gedetailleerde datamodel.

Naast deze basisconcepten worden , zeker bij omvangrijke domeinen, een aantal extra concepten toegevoegd, zoals:

- **Synoniemen of aliassen**, begrippen met dezelfde betekenis binnen het domein maar met een ander begrip
- **Generalisaties of Broader terms**, met name binnen begrippenboom veel toegepast, hiermee wordt hiërarchie aangebracht in de begrippen. Bijvoorbeeld Vervoermiddel – Fiets – Racefiets is een voorbeeld van een hiërarchie waarbij vervoermiddel een generalisatie is van een fiets.
- **Specialisatie of Narrower term**, vergelijkbaar met een generalisatie alleen dan in de richting van de specialisatie. Dus fiets is een specialisatie van vervoermiddel
- **Relatie of related term**, een associatie tussen twee begrippen die aan elkaar gerelateerd zijn. Bijvoorbeeld doordat zij hetzelfde begrip als generalisatie hebben. Bijvoorbeeld Mountainbike is gerelateerd aan Racefiets.

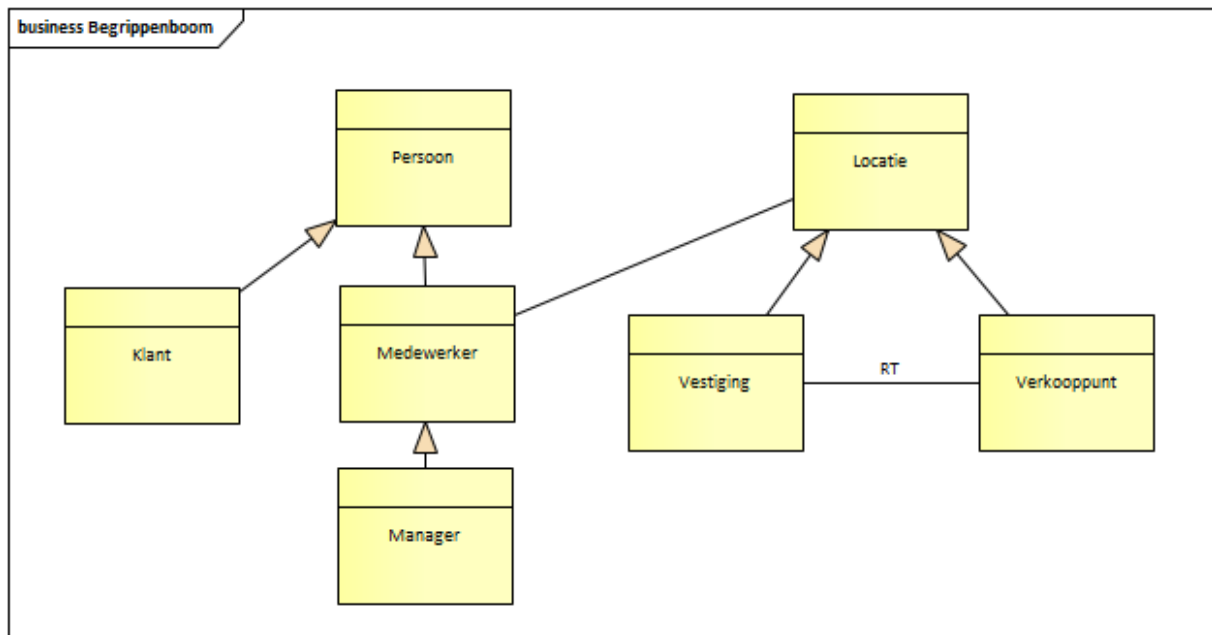
Notatie

Voor de notatie van de begrippenlijst is relatief eenvoudig. Het is veelal een overzicht (op alfabetische volgorde) van de begrippen en hun definitie. Onderstaande afbeelding toont hiervan een voorbeeld.

Name	Alias
Type: Class	
Cursus	
Dit is het meer vaste en inhoudelijke gedeelte van een trainingsaanbod. Het gaat om de elementen die niet veranderen als een training in een andere periode gegeven wordt. Denk hierbij aan inhoud en doel van een cursus	
Training	
Training is de daadwerkelijke bijeenkomst waarin de training gegeven wordt. De attributen hebben daarom met name betrekking op de fysieke aspecten van de training	
Docent	Trainer
Docent is de persoon die de training verzorgt of de cursus ontwikkeld, dit kunnen verschillende personen zijn. Een docent kan zowel TFG maat zijn of van buiten aangetrokken worden.	

In het voorbeeld zie je de begrippen met daaronder kort de definitie die bij het begrip hoort, er is bij een begrip een alias, een docent is in dit domein hetzelfde als een trainer.

Voor begrippenbomen kunnen bomen als modelleerwijze gekozen worden, ook is het mogelijk om gebruik te maken van een eenvoudig grafen model, waarbij de begrippen met gekwalificeerde associaties gekoppeld zijn. In onderstaande voorbeeld zie je hiervan een voorbeeld. De associatie met een driehoekje is een generalisatie – specialisatie associatie, de gerelateerde termen hebben geen extra symbool op de associatie maar de rolnaam RT.



Kenmerken

De begrippenlijst en de begrippenboom hebben de volgende kenmerken:

- Krachtige notatiewijze waarin begrippen op eenvoudige wijze aan elkaar gerelateerd kunnen worden
- Eenvoudig toepasbaar bij gebruik voor stakeholders zonder modellerervaring
- Kan goed gebruikt worden in interactieve workshops
- Toepasbaar op hoge abstractie niveaus, voornamelijk conceptueel
- Goed eerste startpunt bij een top down benadering van een objectmodel
- Goed model om discussie op gang te brengen tussen domeinexperts
- Belangrijk hulpmiddel bij het opstellen van datamodellen bij ketenintegraties.
- Hiërarchieën kunnen complex worden als er veel generalisaties worden gebruikt

Gebruikstoepassingen

Begrippenbomen worden vooral toegepast op een hoog abstractieniveau van datamodellering. Het biedt een goed startpunt voor het in kaart brengen van het gegevensdomein. In complexe domeinen is de begrippenboom een goed startpunt om te komen tot een gezamenlijk domeinmodel waarbij de begrippen de hoogste hiërarchie omvatten.

Houdt er rekening mee dat ondanks de eenvoud van de notatiewijze het opstellen van een begrippenboom een complex traject kan zijn, zeker bij een complex domein of binnen een organisatiecontext waar rond datamodellering weinig volwassenheid is.

Gerelateerde notatiewijzen

Gezien het hoge abstractieniveau van de begrippenlijst en begrippenboom zijn er weinig modellerwijzen die overeenkomsten hebben. Binnen ArchiMate kan met behulp van business objects een begrippenboom opgesteld worden, deze objecten worden vervolgens als koppelpunt naar bijvoorbeeld bedrijfsprocessen of -functies ingezet. Dit wordt nader uitgewerkt binnen de ArchiMate datamodellering whitepapers.

Ook wordt de begrippenboom veelvuldig gecombineerd met een UML klasse diagram. Hiermee ontstaat een hybride datamodel waarbij de begrippenboom als startpunt dient op een hoog abstractieniveau en de UML klassen de detaillering van deze begrippen uitwerken in de klassen, attributen en associaties.

Tooling

Voor de begrippenboom en lijst zijn verschillende tools aanwezig:

- MS-Visio
- Powerpoint
- Excel en Word (voor de boomnotatie)
- Sparx Enterprise Architect
- PowerDesigner
- ERWin
- BizzDesign
- ADO-IT
- ARIS

Evaluatie

De begrippenlijst en begrippenboom zijn krachtige notatiewijzen voor het op hoog abstractie niveau in kaart brengen van begrippen en hun onderlinge relaties. De notatie is eenvoudig en kan daardoor ingezet worden bij stakeholders met weinig modelleerervaring en weinig affiniteit met ICT.

De notatiewijze is eenvoudig echter het opstellen van een begrippenboom kan een langdurig en complex traject zijn als het domein complex is.

Gezien de eenvoud van de notatie is er voldoende tooling aanwezig, van eenvoudige kantoorautomatisering tot geavanceerde architectuur tools.

Over de auteur



Bert Dingemans is trainer op het vlak van data architectuur, data management en Big Data. Hij heeft een passie voor modelleren, modelleertools en het effectief inzetten van geautomatiseerde hulpmiddelen om modellen effectief in te zetten in de praktijk. Bert is te bereiken via bert@interactory.nl