

# Geen goede beslissing zonder integratie BI en CI

## Hoe integreer je BI en CI tot EI

Een aanvaller sprint naar het doel van de tegenstander nadat een medespeler een mooie backspin stiftbal over de centrale verdediging heeft geplaatst. Deze bal valt precies dood in het midden tussen de stilstaande keeper en de snel naderende aanvaller. De aanvaller weet alleen dat hij de 100 meter kan lopen in 12 seconden (BI). Vlak naast hem sprint een verdediger mee die weet dat hijzelf de 100 meter kan lopen in 13 seconden. Maar de verdediger weet ook dat de aanvaller de 100 meter kan lopen in 12 seconden (BI+CI). De verdediger weet dus dat hij het sprintduel gaat verliezen en besluit om de aanvaller even vast te houden aan zijn shirtje om uiteindelijk middels een correcte sliding de bal net voor de voeten van de aanvaller te ontnemen.

**Michael Doves**

---

### **BI + CI = Enterprise Intelligence**

Globaal gezien kan Business Intelligence (BI) worden beschouwd als de informatievoorziening over de interne processen in een organisatie. Daarentegen, kan Competitive Intelligence (CI) worden beschouwd als de informatievoorziening over de externe processen die effect kunnen hebben op de organisatie. CI wordt ook wel de radarfunctie van een organisatie genoemd. In veel organisaties zijn zowel BI en/of CI functies aanwezig, maar zeer zelden zijn deze functies geïntegreerd in een onafhankelijke afdeling tot Enterprise Intelligence (EI). Met

---

## De integratie van BI en CI leidt vaak tot betere inzichten

---

EI is het mogelijk een meer holistische bundel van informatie te creëren, waarmee een organisatie een beter gefundeerde beslissing kan nemen. De intro is een analogie van twee concurrerende organisaties, waarbij de organisatie die BI en CI integreert duidelijk een kennisvoordeel heeft.

Als je BI en CI integreert leidt dat vaak tot betere inzichten. Deze betere inzichten kunnen leiden tot andere beslissingen. Stel dat de aanvaller ook BI en CI had geïntegreerd.

De aanvaller had dan geweten dat zijn snelheidsvoordeel van 1 seconde niet genoeg zou zijn om makkelijk van de verdediger weg te lopen. Een betere beslissing was geweest om eerst met een flinke schouderduw zijn tegenstander weg te zetten tot 50 cm afstand om vervolgens de sprint in te zetten. De kans was groot dat de verdediger zijn shirtje dan niet meer kon vastpakken.

Het integreren van BI en CI tot EI is makkelijker gezegd dan gedaan. Om EI te bewerkstelligen dien je de ongestructureerde wereld van CI te combineren met de gestructureerde wereld van BI. CI informatie wordt veelal vergaard uit niet-gestructureerde bronnen als webpagina's, e-mail, video, audio, RSS feeds, tekstdocumenten, geruchten uit de markt en kennis uit de hoofden van experts en lobbyisten. Uit het CI proces kunnen zachte KPI's zoals Klantsentiment, Concurrentieactiviteit, Partnership Ratio, Voice of the Customer, Social Network Size of zelfs Weak Signals worden gerealiseerd. Wat is de waarde van deze door CI opgeleverde zachte KPI's?

### **Zachte KPI's versus Harde KPI's**

In figuur 1 is een zogenaamd Metrics Continuüm weergegeven. De harde indicatoren als Winst, Kosten en Ziekteverzuim zijn typische indicatoren die bij BI processen worden gerealiseerd. Deze indicatoren worden beschouwd als *lagging*. Lagging indicatoren zijn niet echt voorspellend, maar zeggen iets over de huidige stand van de or-



**Figuur 1: Metrics Continuüm**

ganisatie. *Leading* indicatoren zijn wel voorspellend van aard en betreffen geen absolute harde cijfers, maar eerder relatieve en fuzzy classificaties. Dat is ook meteen het probleem van deze indicatoren; veel mensen kunnen niet goed omgaan met fuzzy indicatoren, omdat deze moeilijk te controleren zijn. Het gebruik van dit soort indicatoren wordt dan ook nog niet veelvuldig toegepast.

Er is een stroming van geleerden die denkt dat voorspellen op basis van lagging indicatoren met behulp van regressietechnieken alleen een goede voorspelling kan opleveren indien de wereld niet aan verandering onderhevig zou zijn. Ik kan het niet bewijzen, maar mijn intuïtie zegt dat deze geleerden het bij het juiste eind hebben. Mijns inziens zou voorspellen een beter resultaat opleveren, wanneer wordt gezocht naar een causaal verband tussen leading indicatoren en lagging indicatoren. Indien een causaal verband wordt gevonden, dan zou dat verband moeten worden meegenomen in het proces van voorspellen. Maar om dat te kunnen doen, zal een organisatie toch eerst de uitdaging van het creëren en monitoren van zachte KPI's moeten aangaan. Hoe kan een zachte KPI door middel van CI worden gerealiseerd?

CI informatie is niet zonder meer te combineren met informatie uit de gestructureerde interne bronnen die tabelgewijs in databases zijn opgeslagen en zijn ontsloten in een data warehouse. De kennis en kunde die hiervoor nodig zijn: business kennis, markt kennis, creativiteit,

taalkunde en ICT kennis.

Om CI informatie bruikbaar te maken voor integratie in BI systemen zal de informatie op een gestructureerde manier moeten worden aangeleverd. Een manier om gestructureerde informatie te destilleren uit een grote berg ongestructureerde data, is door gebruik te maken van 'Ontology Based Text Searching' (OBTS).

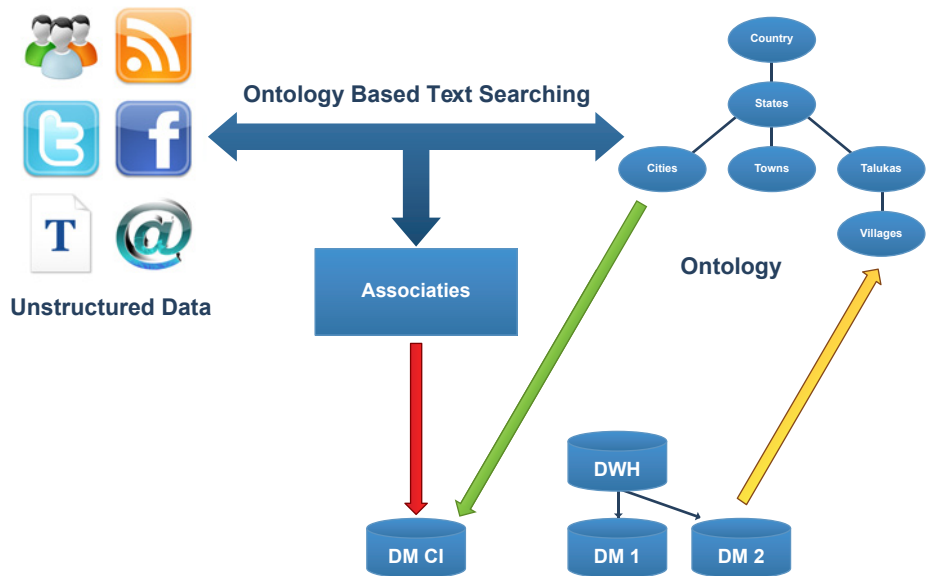
**Associaties**

In figuur 2 is een versimpelde weergave van een mogelijke integratie tussen BI en CI weergegeven. Om OBTS mogelijk te maken, dienen ongestructureerde bronnen als bijvoorbeeld geruchten van de werknemers, Twitter, Facebook, E-mail, RSS feeds en webpagina's aangeleverd te worden aan een OBTS engine. Dit kan zowel handmatig als automatisch gebeuren. Tevens dient een toepasselijke ontologie te worden gerealiseerd, bestaande uit business entiteiten als bijvoorbeeld Concurrenten, Medewerkers, Producten, Trends en Klanten en hun onderlinge relaties. Een dergelijke ontologie schetst een gezamenlijk begrippenkader van het competitieve landschap van een organisatie.

Door het scannen van de ongestructureerde bronnen op basis van de betreffende ontologie kunnen *associaties* tussen de entiteiten en instanties van deze entiteiten worden gevonden en vastgelegd. Associaties zijn verbanden tussen instanties van de entiteiten. Bijvoorbeeld ►►

'Concurrent X werkt samen met Concurrent Y', 'Concurrent X heeft contact met Medewerker Y' of 'Medewerker X schrijft een artikel over Trend Y'.

De associates worden opgeslagen in een relationele tabel van een database. Deze tabel kan dienen als basis voor een feitentabel binnen een dimensioneel model. Het dimensioneel model kan met behulp van bestaande ETL (Extraction, Transformation and Load) technieken worden geladen. Over het algemeen zal gelden dat de business entiteiten in de ontologie de dimensietabellen vormen en de associates de feitentabel. Dit is tevens weergegeven door de groene en rode pijl in figuur 2.



Figuur 2: BI en CI integratieproces

### Weak Signal detection een utopie?

Een ontologie moet natuurlijk onderhouden worden en in veel gevallen dient dat handmatig te gebeuren. Business entiteit Product kan wellicht vanuit een bestaande Product dimensie uit het reeds bestaande DWH automatisch worden ververst. Indien het onderhoud aan de ontologie volledig geautomatiseerd kan worden, dan zijn we waar we willen zijn. Dit wordt weergegeven door de gele pijl in figuur 2.

Stel dat we automatisch trends kunnen ontdekken en deze kunnen toevoegen als instantie onder business entiteit Trend! Het zou heel mooi zijn als je in staat bent om een ontologie dynamisch bij te kunnen werken, waardoor de echte weak signals in de markt ontdekt kunnen worden.

### Dimensioneel model

De tabellen voor rapportagedoeleinden van een data warehouse kunnen we verdelen we in feiten- en dimensietabellen. Feitentabellen bevatten de cijfers, zoals het aantal verkochte eenheden, het aantal uren ziekteverzuim en het aantal dagen met zonnenschijn. De primaire sleutels van deze tabellen worden gevormd door de begrippen waarvan deze feiten afhankelijk zijn. Bijv. het aantal verkochte eenheden is afhankelijk van producten, klanten en winkels. Dit noemen we de dimensies. En de dimensies worden beschreven in - je begrijpt het al - de dimensietabellen. Een model met feitentabellen en dimensietabellen wordt ook wel een dimensioneel- of stermodel genoemd (zie figuur 3).

Deze zogenaamde weak signal detection kan een zeer groot competitief voordeel opleveren door een ramp te voorkomen of een opportunity als eerste te zien.

Niet alles kan eenvoudig worden geautomatiseerd. Soms is CI hard werken en dienen bijv. 2000 LinkedIn profielen te worden gekopieerd en 1000 gevonden CV's te worden ingelezen om een beeld te krijgen van de expertise van de medewerkers van de concurrent. Maar het resultaat van de informatie is holistisch en veel meer op feiten gebaseerd dan wanneer wordt gestuurd op onderbuik gevoel. Er is een mooi gezegde van de heer W. Edwards Deming die zei: "In god we trust, all others bring data".

De kunst is natuurlijk om deze exercitie niet eenmalig te doen, maar er een continu proces van te maken in de organisatie. Het competitieve landschap is natuurlijk maar één ontologie. Men kan natuurlijk ook hele andere ontologieën maken door bijvoorbeeld een 'lead generatie' ontologie op te stellen en de buitenwereld te scannen op leads en prospects. Een ontologie op patentanalyse, een ontologie voor macro-economie, een ontologie voor klantsentiment, een ontologie voor regeringsbesluiten en ga zo maar door zijn ook mogelijk. Op basis van associatie-analyse kunnen de zogenaamde zachte KPI's, die als leading indicators worden beschouwd, in het Metrics Continuüm vorm worden gegeven. Laten deze zachte KPI's nou eens een voorspeller zijn voor de harde KPI's als omzet, kosten en winst. Zou je dan geen mooie forecast kunnen maken waarop je beslissingen kan nemen? ■



**Michael Doves** is Enterprise Intelligence consultant bij adviesorganisatie DIKW Consulting. E-mail: michael.doves@dikw.com

## Praktijkvoorbeeld

Stel er is een fictieve consulting organisatie TestAdvies B.V. die antwoorden wil op de volgende vragen:

- Hoe groot is de Nederlandse test consultancy markt in zowel het aantal consultants als de geschatte omzet?
- Wie zijn de belangrijkste spelers en hoe groot zijn deze?
- Wat is het marktaandeel van TestAdvies B.V.?
- Welke expertises hebben de concurrenten van TestAdvies B.V.?
- Welke klanten hebben deze concurrenten?
- Wat zijn opkomende expertises oftewel trends?
- Wie zijn de *key players* in het management van onze concurrenten?

Door een ontologie te maken van het competitieve landschap van TestAdvies B.V. waarin taxonomieën van entiteiten als Concurrenten, Klanten, Expertises, Personen en Trends zijn opgenomen, kan gestructureerd worden gezocht naar bruikbare associaties in de ongestructureerde data. Een dergelijke ontologie wordt opgeslagen in een database. Met behulp van deze ontologie kan naar verschillende associaties worden gezocht tussen alle entiteiten in de ontologie.

Bijvoorbeeld het scannen van het LinkedIn profiel van Jan Jansen leidt tot de volgende associaties:

- Persoon 'Jan Jansen' is medewerker bij Concurrent 'Testimonial' (AssociatieType 1)
- Persoon 'Jan Jansen' heeft Expertise 'Test Management' (AssociatieType 2)
- Persoon 'Jan Jansen' werkt bij Klant 'BankiLandski' (AssociatieType 3)
- Persoon 'Jan Jansen' werkt aan Trend 'Testen In de Cloud' (AssociatieType 4)

Een RSS feed op het Financieel Dagblad leidt tot de volgende associaties:

- Concurrent 'Testimonial' heeft een partnership met Concurrent 'LogicalTesting' (AssociatieType 5)
- Concurrent 'Testimonial' heeft een opdracht binnengehaald bij Klant 'InsuranceCar' (AssociatieType 6)
- Concurrent 'Testimonial' doet onderzoek naar Trend 'Automatisch testen voor Data warehousing' (AssociatieType 4)

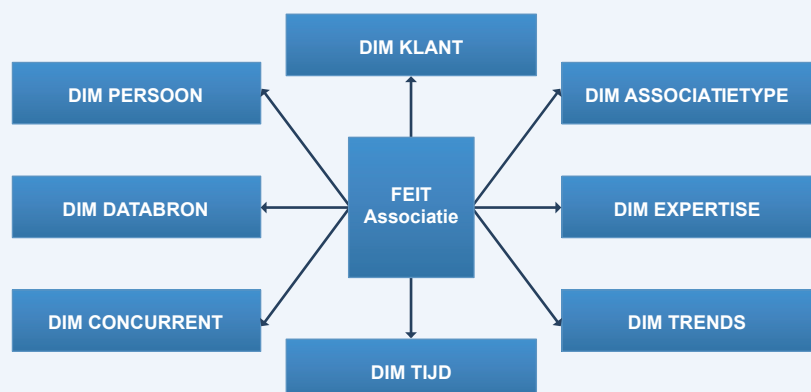
Met een beetje creativiteit kan door middel van het scannen van onge-

structureerde bronnen, op basis van kernwoorden of kern frases uit een gedefinieerde ontologie, heel veel bruikbare associaties worden vastgelegd. Deze associaties met de bijbehorende ontologie wordt vertaald naar een sterdiagram in het BI domein.

De feitentabel bestaat uit de associaties en de dimensies bestaan uit de entiteiten van de ontologie. De dimensies bevatten een hiërarchie wanneer de entiteiten taxonomieën hebben. Het sterdiagram ziet er dan uit als geïllustreerd in figuur 3. Op basis van dit sterdiagram kan dan ook een multidimensionale database worden gedefinieerd. Door te slicen, dicen en drillen kunnen hele leuke inzichten in het competitieve landschap van TestAdvies B.V. worden opgedaan.

Met behulp van een aantal aannames gebaseerd op BI informatie over je eigen omzet per consultant, bijvoorbeeld een consultant draait gemiddeld 120.000 euro omzet per jaar, kunnen dan hele mooie schattingen worden gemaakt over de grootte van de markt en de omzet per Concurrent en Klant. Door taxonomieën of hiërarchieën in de Klanten en Concurrenten op te nemen als branche en regio, kunnen ook nog mooie markt cijfers per branche en regio worden gepresenteerd.

De feitentabel bevat per record vaak alleen verwijzingen naar slechts twee dimensies. De meetwaarde is eigenlijk alleen een 1. Men noemt dat ook wel een feitenloze feitentabel. Er is wel altijd een verwijzing naar de Meta Dimensies als Dimensie Associatietype, Dimensie Databron en Dimensie Tijd. Vanuit deze informatie kan een zachte KPI genaamd Klant Trend Index worden gedestilleerd. Hieruit kan worden gehaald welke trends daadwerkelijk doorzetten en welke alweer afnemen.



Figuur 3: DM CI voorbeeld sterdiagram