



Gartner beloont Utrechts WCC met plaats in magisch kwadrant

Sterke positie in HR en identity matching

René Rippen

In het magisch kwadrant van marktonderzoeker Gartner dat de aanbieders van toegang tot informatie in kaart brengt, verschijnt in oktober plotseling een klein Nederlands bedrijf: WCC Smart Search & Match. Reden genoeg dus voor een gesprek met de oprichter en het management.

Het zal niet vaak gebeuren dat we voor een dergelijk bezoek niet verder hoeven te rijden dan industrieterrein Lage Weide bij Utrecht. WCC heeft zijn ontstaan te danken aan de frustrerende zoektocht naar een huis, vertelt oprichter en CEO Peter Went. "Als een echte techneut had ik me goed voorbereid met een *request for proposal* op een A4-tje. Daarmee ben ik naar een makelaar gegaan. Die reageerde met: 'Nee, mijnheer, zo werken wij niet. Ik wil eerst weten wat voor bedrag u wilt besteden, of u een huis wilt of een flat, op welke plek u wilt wonen'. Terwijl ik toch dacht dat acquisitie en het koppelen van mij aan een huis zijn kernactiviteit was. Iets dergelijks overkwam een vriendin van me bij het zoeken naar een baan. Een zoektocht door de vacaturebank leverde niets op, terwijl een advertentie in de lokale krant van die week de perfecte baan bleek op te leveren.

Geld was in beide gevallen het probleem niet: alles was geautomatiseerd, behalve de betreffende kernactiviteit: matching.

Ik ben van origine een database-man met DB2-ervaring bij Time Life en Shell en heb in het weekeinde en 's avonds na zitten denken over dit probleem. Mensen doen heel vaak aan matching: zo komen ze aan een auto, een huis, een levenspartner, een baan. Ze komen niet tot hun keus via Boolean logica; die kan niet overweg met eigenschappen als kleur, snelheid, degelijkheid, zuinigheid. In de praktijk gaat het om grijstinten en databases werken niet met nuances, prioriteiten, grijstinten. Daarvoor is geen plaats in Boolean *true or false*."

Niet relationeel

Went begint in 1996 zijn eigen bedrijf. Twee maanden later ontmoet hij Mark Wegman, de CTO van WCC. Die heeft dan net de middelbare school afgerond en er al een aantal projecten als programmeur opzitten. Wegman adviseert Went af te stappen van het idee dat matching op grote schaal via een relationele database een haalbare kaart is. Went: "Daar had Mark groot gelijk in,

maar de markt opgaan met een eigen database is een flinke handicap. We hadden dus een commerciële reden om voor een commerciële database te kiezen. Uiteindelijk is dat SQL Server van Microsoft geworden, omdat die de grootste acceptatie in de markt heeft."

Went en Wegman komen in contact met een van de voorlopers van PinkRoccade, waar men een integraal systeem bouwt voor Arbeidsvoorziening van het ministerie van Sociale Zaken en Werkgelegenheid. Went: "Daar hebben we een aantal matching benchmarks gedraaid. Al snel moesten we concluderen dat het met een relationele database als engine met 600.000 werklozen absoluut niet zou werken. Zonder ranking komt matchen neer op een *tablespace scan* lineair door alle records heen. Dat is met zoveel records onaanvaardbaar traag. Wij konden het met de ranking van onze zoek- en match-engine Elise veel en veel sneller."

De beursgang van Google heeft er voor gezorgd dat zoeken van informatie een 'hot' onderwerp is geworden

Dat resulteert op 18 september 1997 in het eerste contract voor de jonge starter met Everest¹, dochter van het latere PinkRoccade voor de bouw van het systeem voor Arbeidsvoorziening. Went: "Na het binnenhalen van die opdracht dachten we: als zo'n overheidsorganisatie met een enigszins stoffig imago het met ons en onze technologie ziet zitten, dan gaan we succesvol worden. Dan hebben we ook marketing en sales nodig. In 1999 komt de vaart er in en halen we met Vedior, Randstad en Adecco drie grote klanten binnen in de HR-sector, onze historische markt."

Solide bedrijf

WCC Smart Search & Match voert een conservatief beleid, zegt Went. "En ook onze Raad van Commissarissen stimuleert dat; we zijn een solide bedrijf met klanten onder grote internationale bedrijven. En we zitten in het kernproces van die ondernemingen. Er zit geen extern geld in de onderneming van *venture capitalists* en dus is er ook geen externe druk om naar de beurs te gaan. Onderhoud van geleverde systemen en additionele licenties zorgen voor een stevige basis in onze omzet en voor een solide groei. We zijn jarenlang per jaar voor één prospect gegaan en hebben die tot klant gemaakt. Maar de markt is aan het veranderen; de beursgang van Google heeft er voor gezorgd dat zoeken van informatie een 'hot' onderwerp is geworden. Er zijn ook nieuwe concurrenten opgedoken in de markt, zoals Autonomy en Fast. Bedrijven zien de waarde van zoektechnologie nu in en bovendien is de economie flink aangetrokken. Met andere woorden: WCC moet meer naar buiten treden. We hebben onze sales en marketing in Europa versterkt met Johan Morsink en Alexandra Buytendijk. En tot onze grote blijdschap blijkt dat te werken, gezien onze entree in het magische vierkant van Gartner. Maar we zijn maar met 35 man en kunnen niet alles tegelijk: geografische expansie en uitbreiding naar andere bedrijfstakken." De basis van WCC ligt in de HR: arbeidsvoorziening, uitzendbureau's en job-boards op internet. Sinds een jaar of twee is het bedrijf ook actief op de markt van identity matching en content & records management. Het vaststellen van de persoonsidentiteit is sinds de aanslagen van 11 september 2001 een groeimarkt geworden. "Daarin zijn op instigatie van en gefinancierd door nationale overheden heel veel projecten gestart, met name bij politiediensten en grensbewaking."

Identiteit van personen

Persoonskenmerken als naam, geboortedatum, DNA, gezicht, stem, iris en vingerafdruk worden gebruikt om de identiteit van personen vast te stellen. WCC is in die wereld terechtgekomen door een succesvol traject met Accenture bij de Arbeidsvoorziening in Duitsland. Dat leverde zoveel positieve respons op dat Accenture besloot WCC in te schakelen bij een enorm project dat de consultancy-onderneming in 2004 binnensleepte bij US Visit, de grensbewaking van de VS. "Ze zoeken daar zoals elke Amerika-reiziger weet op gezicht en vingerafdruk. Bij uitstek geschikt voor onze ranking-technologie, want perfecte Boolean-true matches bestaan er niet bij biometrische data."

Met vier Gigabyte aan geheugen per PC heeft de database aan 250 PC's voldoende om de database in memory op te slaan

Er zijn thans 64 miljoen personen opgenomen in dit register; dat moet uitgroeien tot 500 miljoen personen. Het aantal vingerafdrukken gaat binnen afzienbare tijd omhoog van de huidige twee wijsvingers naar alle tien vingers. "Als de database helemaal gevuld is, zitten er 5 miljard vingerafdrukken in, ongeveer 1 Terabyte aan data. Met vier Gigabyte aan geheugen per PC heeft de database aan 250 PC's voldoende om de database in memory op te slaan." WCC maakt ook deel uit van het Accenture-consortium dat tracht de opdracht voor een Europees

Fuzzy en schaalbaar

'Matchen staat gelijk aan ranken' is het adagium bij WCC. CTO Mark Wegman: "Bij een HR-systeem wordt Elise, onze zoek- en match engine gevoed met data over de sollicitant: persoonlijke gegevens, werkervaring, gewenste baan, salariseisen en dergelijke. Elk van die sectoren krijgt een score. Daar rolt een totaalscore uit, die in een ranglijst van sollicitanten uitmondt. Die ranglijst wordt gebruikt bij de matching aan het profiel van een vacature. Ook die wordt in een ranglijst gezet. De score kent dus twee dimensies: 'Ik vind de baan leuk' en 'De baan vindt mij leuk.'" Elke request begint op nul; vooraf vindt geen indexering of optimalisatie plaats. Wegman: "Daar is de architectuur op ingesteld. We hebben alles zelf ontwikkeld in Elise: de database engine, geschreven voor de opslag in het geheugen, de match engine, de ondersteuning van het gedistribueerde platform. Aan het feit dat we de expertise in huis hebben en alles zelf hebben gebouwd, danken we de enorme snelheid van ons platform." WCC heeft al vroeg in zijn bestaan een tweetal technologische keuzes gemaakt die heel profijtelijk zijn gebleken: de database in het geheugen laten draaien en gedistribueerd over verschillende computers. Went noemt het onderkoeld "technische risicofactoren" en erkent dat WCC

daarmee risico's heeft gelopen. "We wilden de ultieme technische keuze maken en konden dat doen omdat we niet onder druk stonden van *venture capitalists* in onze onderneming. De erosie in de prijs van geheugenchips hebben we aan zien komen, net als de toenemende rekenkracht van PC's. Vandaar de vroegtijdige keus voor een gedistribueerde en in memory draaiende database.

Het risico is dat je met de technologische keuzes te vroeg in de markt komt. Breedschalige marketing werkt niet; je moet in het begin een heleboel evangelisatiewerk doen", erkent Went. "Je klanten moeten een prangend probleem hebben en tegelijk open staan voor een innovatieve oplossing. Maar als je bij een project bij de Arbeidsvoorziening in Groot-Brittannië de Ingres databases in vier rekencentra kunt repliceren naar een paar PC's met onze Elise software en dan met 60.000 consultants *subsecond* kunt zoeken in subseconds, dan weet je dat je goed gekozen hebt. Na wat we sindsdien van de architectuur van Amazon en Google weten is die keus thans begrijpelijk, maar dat was hij indertijd allerminst."

Meer informatie is op www.wcc-group.com te vinden.

biometrisch systeem binnen te halen. Naast deze toepassingen uit de publieke sector heeft ook de private sector baat bij de ranking-technologie van WCC. Went: "Bij de Concentra verzekeringsmaatschappij in de VS helpt onze Elise zoek-engine de atientgegevens te vinden van de 17 miljoen ziektenkostenverzekerden, ook als er tikfouten zijn gemaakt bij de invoer of de zoekvragen.

Messenger-data

WCC zit met zijn toepassing in tussen de klassieke SQL-query's van de gestructureerde databases met hun Boolean logica en de tekstgeoriënteerde *enterprise search engines* van Endecca, Fast, Autonomy en Google. "Met wat wij noemen 'universal meaningful search' ontsluiten we zowel gestructureerde informatie uit databases als Messenger-data, spraak, beeldmateriaal en andere ongestructureerde content." Johan Morsink, VP business development EMEA kent daarvan een mooi voorbeeld. "Een Nederlandse bankinstelling regelt zijn autorisatie en authenticatie via een kopie van paspoort of rijbewijs, dat dan naar het hoofdkantoor wordt opgestuurd. Zoeken in die gescande OCR-bitmaps bleek een te zware opgave voor de oorspronkelijke zoek-engine. Nu hebben wij die opdracht gekregen, puur omdat we het sneller konden, met minder fouten en met minder hardware." WCC Smart Search & Match verkoopt zijn systemen direct aan de klant. Nu er meer klanten komen en er vaker indirect via partners

Gartner over WCC

"De sterkste markten voor WCC zijn het samenbrengen van vacatures en sollicitanten, identity matching en e-commerce. WCC is een kleine aanbieder met zijn hoofdzetel in Nederland en zijn centrale technologie verbetert de respons op gestructureerde query's, door overeenkomsten in data records zo te interpreteren dat de zoeklogica in staat wordt gesteld tot nuttige antwoorden te komen. De software komt het best tot zijn recht met gestructureerde data, maar verschaft ook inzicht in data uit documenten, multimediale data en biometrische data."

Uit: Gartner, Magic Quadrant for Information Access Technology, oktober 2006.

of de grote systeemintegrators wordt verkocht, wordt de nauwe band met de klant ongetwijfeld losser, zegt Went. "Maar één ding blijft: de klanten begrijpen hoe belangrijk een goed werkende match-engine voor hen is."

Noot

1. Zie ook *Business Process Magazine* nr 8/2006 voor een interview met Leo Hermans van Everest.

René Rippen is freelance journalist.

Update

Business Objects challenger in Magic Quadrant data-integratie

Business Objects is door Gartner als challenger opgenomen in het Gartner Data Integration Tools Magic Quadrant, gepubliceerd op 22 november 2006. Business Objects biedt data-integratie, metadata management en datakwaliteitsoplossingen als onderdeel van een uitgebreide productlijn die bedrijven ondersteunt op het vlak van enterprise information management (EIM). De datakwaliteitsoplossingen van Business Objects integreren en verbeteren informatie op een snelle en flexibele manier, zodanig dat een organisatie beslissingen kan nemen op basis van betrouwbare gegevens. Eerder dit jaar nam Business Objects het bedrijf Firstlogic over om zijn data-integratie portefeuille uit te breiden en zich duidelijker te profileren binnen deze

markt. In 2006 stegen de licentie-inkomsten uit BO's EIM-producten tot meer dan 31 miljoen dollar in de eerste drie kwartalen, een stijging van bijna 60% ten opzichte van dezelfde periode in 2005. www.businessobjects.com

Netezza en IBM leveren verbeterde gegevens-integratiefunctie

Ondersteuning door Netezza's high-speed dataloader maakt het gebruikers van IBM Information Server mogelijk de prestatie van het NPS-systeem volledig door te voeren in hun gegevensintegratie. Door deze integratie kunnen klanten IBM Information Server naadloos doorvoeren in het Netezza datawarehouse-platform en tegelijkertijd de prestatie maximaliseren. IBM Information Server ondersteunt Netezza's high-speed data loader, die momenteel een gegevensoverdrachts-

snelheid van 500 Gigabyte per uur bereikt, en stelt IBM Information Server-klanten in staat een zeer hoge verwerkingscapaciteit te bereiken bij het laden van gegevens in het NPS-systeem. Het Performance Server-systeem van Netezza is een datawarehouse-platform dat speciaal ontwikkeld is voor sneller en goedkoper analyseren van Terabytes aan gedetailleerde gegevens, dan bestaande datawarehouse-opties. Het NPS-systeem slaat Terabytes aan dossiers op en filtert en verwerkt deze in één eenheid, waarbij alleen de relevante informatie voor elke vraag wordt geanalyseerd. Netezza heeft CPU-kracht naast de gegevens geplaatst, waardoor het NPS-systeem razendsnel processen uitvoert waarmee de meeste datawarehouse-systemen uren of zelfs dagen mee bezig zouden zijn en waardoor een spectaculaire toename van productiviteit binnen de organisatie mogelijk is. www.netezza.com