

Alkmaars bedrijf biedt meer dan louter hulp bij textmining

# Tarchon ondersteunt groeiend leger kenniswerkers

Paul van der Linden en Izabella Smits-Pukrop

**W**ork at the speed of your thoughts' is het motto van Tarchon. Het Nederlandse bedrijf, op 1 april 2001 opgericht door Vincent van den Brekel en Henk Alles, levert software die informatie uit willekeurige bronnen ontsluit en op intuïtieve wijze aanbiedt aan kenniswerkers. CTO Henk Alles legt uit dat het jonge bedrijf zich momenteel bevindt in 'fase één' van zijn bestaan: "Tarchon moet geld opleveren. Hiervoor hebben we negen maanden."

Alles is de chief technical officer (CTO) van Tarchon, Vincent van den Brekel zwaait de directeurscepter. Het startkapitaal is afkomstig uit Duitsland. We hebben het over een *core technology company*. Alle energie wil men richten op het leveren van goede technologische producten die kenniswerkers daadwerkelijk ondersteunen bij hun dagelijkse werkzaamheden. De implementatie van de software en het inbedden van de bijbehorende werkwijze in de organisatie laat Alles graag over aan de consultancybureaus. "Die beschikken namelijk over de materiedeskundigheid die in een specifieke situatie nodig is", aldus Alles. Hij geeft aan dat er nog veel aan evangelisatie gedaan moet worden; tegen het einde van 2001 had Tarchon zes klanten, waaronder Udink & De Jong, Andastra en First Dutch Capital.

## BI EN CI

In het tweede kwartaal van dit jaar denkt Alles de tweede fase te kunnen ingaan, waarin nieuwe producten worden gelanceerd. Wat zit er in de pijplijn? Dezer dagen komt een product voor kennisdeling op de markt (codenaam: Triton), een serverproduct. Gebruikers kunnen zich abonneren om toegang te krijgen tot de aanwezige kennis door gebruik te maken van een 'Subscribe me'-optie. Daarnaast is er een BI-tool, Beowulf. Dit werkt met profielen en zoekt actief naar concurrentiegegevens en dergelijke. Deze software valt hiermee zowel in de markt van business intelligence als competitive intelligence. Dan is er Akelei, een *local area search engine* voor internetsites, een derde product waaraan Tarchon momenteel werkt. En ten slotte wordt een enterprise information

portal builder aangeboden. Alle producten hebben direct te maken met het ontsluiten en beschikbaar stellen van informatie en kennis.

## GLASPLATEN

Henk Alles refereert aan de uitspraak dat elke zeven maanden een verdubbeling van gegevens plaatsvindt. "Je moet gebruikers helpen om vragen te stellen. In de meeste gevallen zijn ze hiertoe zelf namelijk niet in staat." De genoemde verdubbeling van gegevens helpt hierbij natuurlijk ook niet. Hoe kunnen gebruikers vinden wat zij zoeken als zij geconfronteerd worden met een heterogene

## Goudzoeken in ongestructureerde data

In nummer 7 van vorig jaar heeft DB/M in een thema-artikel aandacht besteed aan textmining. Gartner beschouwt textmining als 'het minder bekende neefje van datamining'. Terwijl datamining is bedoeld om kennis te ontsluiten uit gestructureerde gegevens, zoals relationele databases, richt textmining zich op de groeiende hoeveelheid ongestructureerde data, onder meer uit e-mails, teksten en Powerpoint-presentaties. De markt voor textminingssoftware is momenteel nog erg ondoorzichtig. Dat heeft veel te maken met de relatieve onvolwassenheid van die markt. Leveranciers die claimen tools voor textmining aan te bieden, hebben een zeer uiteenlopende achtergrond: zij bewegen (of bewogen) zich in markten voor systemen voor document management, zoekmachines, scanningproducten en werkgroepssoftware. Betekent dit dat organisaties zowel dataminingsoftware als een textminingtoepassing in huis moeten hebben om al die data betekenisvol te kunnen ontsluiten? Gelukkig niet. In dit artikel wordt ingezoomd op Tarchon, een softwaresuite waarmee informatie uit willekeurige bronnen -ongestructureerd dan wel gestructureerd- kan worden ontsloten en op een intuïtieve wijze aan kenniswerkers wordt aangeboden.

## Fuzzyness en semantische netwerken

Hoe werkt de software van Tarchon? Alles legt het uit aan de hand van twee begrippen:

- *fuzzyness*. De software werkt op een associatieve manier en vooraf. Bij een spellingfout herkent reguliere software het betreffende onderwerp niet. Door gebruik te maken van fuzzyness worden onderwerpen ook in dit geval getraceerd: een zoektocht naar 'Amsterdam' levert ook 'Amsterdan',



### EEN ZOEKTOCHT NAAR "ROMSTERDAM"?

'Amsterdamned' en 'Amsterdma' op. Aard en mate van het fuzzy zijn -wat moet wel, wat niet een treffer geven- kunnen overigens gestuurd worden;

- Tarchons software maakt gebruik van semantische netwerken. Een zoekopdracht naar 'Alkmaar' levert nu niet honderdmaal

de term 'Alkmaar' op, maar slechts eenmaal, met verwijzingen naar andere onderwerpen, zoals 'kaas' of 'Tarchon'.

Beide aspecten resulteren volgens Alles in een betere performance.

Tarchon maakt gebruik van een eigen grafendatabase, een zelf ontwikkelde, semantische database. Henk Alles: "Deze doet ongeveer wat een objectgeoriënteerde database doet. OO databases zijn echter allemaal transactiegeoriënteerd. Ze komen hiermee in de buurt van wat gewenst is, maar bieden ook onnodige functionaliteit, die alleen maar in de weg zit. Daarom hebben we ervoor gekozen zelf een database te maken." Vragen kunnen worden gesteld in een natuurlijke taal. In de database worden niet de verschillende brongegevens opgeslagen, maar alleen de relaties ertussen en de associaties. Tot de functionaliteit van de database behoren dan ook clustering, aggregatie en dergelijke.

Bij het zoeken naar een bepaald onderwerp maakt Tarchon uit de resultaatset één nieuw document. En dat is ook eigenlijk wat je wilt. Je wilt de informatie uit de honderd documenten die de zoektocht heeft opgeleverd, maar je wilt niet de honderd documenten als zodanig. Met behulp van een standaard functie is het nu mogelijk een selectie te maken uit de gegenereerde steekwoorden. Vervolgens kan op basis van de geselecteerde steekwoorden een samenvatting worden gegenereerd. Dit gebeurt op een hoog interactieve manier. De gebruiker kan onder meer aangeven hoe groot de samenvatting moet zijn. De presentatie is ook sterk intuïtief. Steekwoorden en zinsneden zijn grafisch presenteerbaar. Door deze steekwoorden te volgen 'langs hun relatie' ontstaat al door associatie een begrip van wat het geheel aan informatie eigenlijk inhoudt.

databrij? Een correcte en intuïtieve presentatie van de enorme hoeveelheid data stelt een gebruiker in staat om op de juiste vraag te komen. Voor personen met een OLAP-achtergrond een bekend fenomeen. Door presentatie van de gegevens ontstaan al of niet vragen, die de gebruiker van te voren niet kon stellen. Hij wordt geconfronteerd met de gegevens (de resultaten), en louter daarvoor komt een vraag bij hem op.

### Hoe kunnen gebruikers vinden wat zij zoeken in een heterogene databrij?

Dat geldt ook in het geval van Tarchon. Met als verschil dat we het hier niet hebben over OLAP, maar over het veel bredere gebied van kennismanagement. Is het 'werkterrein' van OLAP nog beperkt tot gestructureerde data, in dit geval gaat het ook om de ongestructureerde gegevens.

Tarchon wil kennis tastbaar maken. De database bevat een

semantisch netwerk per document. Henk Alles spreekt in dit verband over "glasplaten". *Profiling* is dan alles wat iemand interessant vindt: "rode glasplaten".

## INFORMATIE EN KENNIS

Geen inspanning vooraf, is het adagium dat Alles aanhangt; dat is simpelweg niet meer economisch verantwoord, gezien de enorme berg informatie waarover we inmiddels beschikken. De markt voor kennismanagement wordt becijferd op 1,8 miljard Amerikaanse dollar, met een jaarlijkse groei van 44 procent. Als bedrijf dat zich uitsluitend bezighoudt met het ontwikkelen van zijn technologie laat Tarchon het aan partners over om de software te implementeren, zoals Cap Gemini Ernst & Young en Origin. In de loop van dit jaar gaat men werken met een dealerkanaal (voor Duitsland en België).

Terwijl Pink Roccade kraait dat informatie alles is, roept Tarchon dat informatie nutteloos is. Organisaties doen volgens Tarchon niets anders dan informatie opslaan en vergaren. Hergebruik is

daarbij het achterliggende doel. Dit voortdurend opslaan van informatie betekent dat elk halfjaar de capaciteit daarvoor moet worden verdubbeld. Maar het ontbreken van effectieve zoekmethoden ontmoedigt medewerkers zoekacties uit te voeren. Zo worden kansen gemist om informatie te transformeren naar kennis. De rentabiliteit van de opgeslagen informatie is hiermee lager dan beoogd en dan mogelijk is.

Is informatie waardeloos, kennis is waardevol. Door gericht en intuïtief te kunnen zoeken in de aanwezige informatie is het mogelijk om vanuit deze informatie vervolgens kennis te genereren. En dat is precies wat de kenniswerker nodig heeft. Hij zoekt informatie, genereert kennis en kan die vervolgens vertalen in acties. De rol die Tarchon hierbij wil spelen is die van aanbieder van een tool voor informatie-extractie en presentatie. Kennis is in de definitie van Tarchon het onthouden van eerder geleerde zaken. Als het gaat om het leren op cognitief gebied vormt kennis het laagste niveau.

**OPBOUW**

Tarchons oplossing bestaat uit vier basisconcepten. Intuïtief is het eerste concept - hetgeen hier zoveel betekent als aansluitend bij het verwachtingspatroon en het delen van kennis door het bouwen van *communities*. Het tweede concept is dat informatie slechts eenmaal is beschreven en opgeslagen in een efficiënt kennissysteem. Op deze manier bestaat er geen noodzaak data te dupliceren en te synchroniseren. Gegevens blijven staan waar ze oorspronkelijk zijn neergezet. Alleen de vindplaatsen, verwijzingen en verbanden worden opgeslagen. Homogeniteit is het derde basisconcept. De heterogene informatiebronnen worden als homogeen gepresenteerd door middel van een *single point of access*. Het vierde concept is flexibiliteit. Tarchon bouwt een virtueel kennisnet-

**Een zoekopdracht levert niet de verfoeide lijst met honderdduizend min of meer relevante hits op, maar een overzicht met de relevante steekwoorden en de relaties daartussen**

werk waarbij verbanden en relaties dynamisch groeien en krimpen.

Tarchon biedt haar software aan in suite, die bestaat uit de volgende onderdelen: Rosetta, Strannix, Nebula, Praetorian en Merlin. Rosetta betreft het verzamelen van de inhoud (*content gathering*). Rosetta doorzoekt de heterogene informatiesystemen op het intranet of Internet en zet alle data om in XML. Strannix vormt de eigenlijke kern van het systeem. Het legt relaties en verbanden tussen de door Rosetta gevonden informatie. Hiermee wordt de informatie in haar context geplaatst. De door de semantische processor Strannix gelegde verbanden worden vervolgens door



**FIGUUR 2: JE WILT DE INFORMATIE UIT DE HONDERDEN DOCUMENTEN DIE DE ZOEKTOCHT HEEFT OPGELEVERD, MAAR NIET AL DIE DOCUMENTEN ALS ZODANIG.**

Nebula opgeslagen. Nebula zorgt daarnaast ook voor de ontsluiting van de opgeslagen informatie. De toegangsrechten voor eindgebruikers worden vastgelegd in Praetorian. Merlin ten slotte is de gebruikersinterface die op een intuïtieve manier toegang verschaft tot de opgeslagen informatie. De gebruikers kunnen iteratief en op basis van hypothesen zoeken in zowel gestructureerde als ongestructureerde gegevens.

**DOELGROEPEN EN CONCURRENTIE**

Wat zijn volgens Henk Alles de concurrenten voor Tarchon? Op de eerste plaats noemt hij de categorie van de *“old fashion information retrieval software”*, waartoe ook Verity behoort. “Een beetje verouderde software, maar wel een concurrent.” Daarnaast noemt hij Inktomi: internationaal gezien wel een bekende speler, maar in Nederland niet erg actief. “Goede technologie, maar niet bezig op kennisgebied”, aldus Alles. De queeste van Tarchon is kennismanagement gemeengoed (een *commodity*) te maken. Het duurste wat Tarchon te bieden heeft, kost dan ook 10.000 euro; van onder af gezien begint het prijslijstje bij 2.499 euro. Tarchon vindt het belangrijk dat de software die ze levert eenvoudig onderhoudbaar en installeerbaar is.

Tarchon mikt op de zogenaamde kennisintensieve organisaties (KIO's). Daarin zijn grote groepen kenniswerkers aanwezig die een groot deel van hun tijd omgaan met informatie. Het destilleren van kennis en het gebruik ervan (omzetten in acties) is hun voornaamste bezigheid. Vaak is 'marketing' de eerste afdeling waar de Tarchon-software wordt gebruikt. Organisaties als grote banken, advocatenkantoren en verzekeringsmaatschappijen zijn KIO's, maar denk bijvoorbeeld ook aan consultancy's en overheden. Tarchon wordt momenteel gebruikt bij The Boston Consulting Group en Hope Investment Bank.



**FIGUUR 3: DOOR STEEKWOORDEN TE VOLGEN 'LANGS HUN RELATIE' ONTSTAAT DOOR ASSOCIATIE AL EEN BEGRIIP VAN WAT HET GEHEEL AAN INFORMATIE EIGENLIJK INHOUDT.**

De software ondersteunt drie talen (Duits, Engels en Nederlands) en herkent er zeventig. Tarchon wordt onder meer gebruikt voor het personificeren van *investment proposals*, het vasthouden van kennis uit e-mails en bijlagen (Udink & De Jong), reisportals voor zakelijke reizigers en voor het aanbieden van packages (Andastra). Doel is om eind 2002 acht, negen of tien klanten te hebben. Alles geeft aan de tijd te willen nemen om zaken goed op te zetten. Van de vijftien medewerkers die de onderneming nu telt nemen zijn er twee bezig met sales. Het leeuwendeel neemt de verdere productontwikkeling voor zijn rekening. Daarnaast is Tarchon gelieerd met een bedrijf in Oekraïne, waardoor desgewenst tot zeventien ontwikkelaars bijgeschakeld kunnen worden.

**TECHNOLOGIE**

De Tarchon-technologie is Windows NT/2000-software en draait in een serveromgeving. Er is nog geen Unix-versie beschikbaar. Tarchon kan zowel gestructureerde als ongestructureerde informatie te ontsluiten, waaronder tekstbestanden, e-mailbestanden, PDF's en html-pagina's en maakt gebruik van een webserver. Uit eerdere projecten is gebleken dat de software binnen een periode van vier weken is te implementeren.

Een licentie wordt berekend per server en is niet gebonden aan het aantal eindgebruikers. Tarchon werkt momenteel samen met Object Technology om het systeem te koppelen aan CRM-systemen, beslissingsondersteunende oplossingen en andere toepassingen.

Hoewel Tarchon pas sinds 1 april 2001 actief is, kan de Alkmaarse onderneming bogen op een lange achtergrond. De kern van het bedrijf, waartoe ook Alles kan worden gerekend, houdt zich sinds meer dan vijftien jaar bezig met intelligente taaltechnologieën. De methoden die worden gebruikt om bruikbare herkenningpatronen te verkrijgen bestaan ruim twintig jaar en zijn terug te voeren op het werk van Russische wetenschappers. Technologie en opbouw van Tarchon vinden hun basis dan ook in de voormalige Sovjetunie. Tot de huidige klantenkring van Tarchon behoren onder andere het Duitse internetbedrijf Andastra en de First Dutch

Capital Group (DCG) uit Amsterdam. De verhouding tussen research en analyse, gemeten in hoeveelheid werktijd, wordt volgens het laatste bedrijf met Tarchon omgedraaid van 80-20 naar 20-80%.

**HOOFDEN**

Tarchon biedt meer dan alleen textminingfunctionaliteit. De geboden software sluit heel goed aan bij de groeiende behoefte om daadwerkelijk eens met kennismanagement een serieuze start te maken. Het onderwerp leeft al langere tijd, maar heeft de afgelopen jaren veel aan kracht gewonnen. De vloedgolf aan gegevens die door de hedendaagse organisaties spoelt, gecombineerd met het feit dat het ware bedrijfskapitaal vaak in de hoofden van medewerkers zit, geeft aan dat kennismanagement een serieuze aangelegenheid moet zijn. Bovendien groeit het leger van kenniswerkers, personen die voor het grootste deel van hun tijd bezig zijn met het zoeken naar en analyseren van gegevens. Dat maakt het belang van goede ondersteuning van deze groep duidelijk.

Tarchon heeft een aantal duidelijke *unique selling points*. In de eerste plaats de manier waarop aan data-ontsluiting wordt gedaan. Er wordt niet met gegevens geschoven. In plaats daarvan worden vindplaatsen en relaties vastgelegd. Alle soorten gestructureerde

*Informatie is waardeloos, kennis is waardevol*

en ongestructureerde gegevens worden daarbij toegankelijk. Een zoekopdracht levert niet de bekende -en verfoeide- lijst met honderdduizend min of meer relevante hits op, maar een overzicht met de relevante steekwoorden en de relaties daartussen. Een samenvatting over een set van documenten kan on-the-fly worden gemaakt. De software werkt daarnaast intuïtief en sluit goed aan op de werkwijze van de hedendaagse kenniswerker. En ten slotte - ook niet geheel onbelangrijk- is de prijsstelling dusdanig, dat laagdrempelig het pad naar kennismanagement ingeslagen kan worden. Geen excuus meer om het uit te stellen. ●

Drs. P.F.H. van der Linden (P.vanderLinden@Synergetics.nl) is principal consultant bij Synergetics, The Management Information Company. Izabella Smits-Pukrop (I.Smits-Pukrop@Synergetics.nl) is consultant bij Synergetics.