



Explicit Content

Als u, net als ondergetekende, in de database wereld thuisloopt dan wil dat nog niet zeggen dat u iets met data heeft. "Data, dat is iets van en voor gebruikers" lijkt impliciet het motto van de database man of vrouw. Wij database experts houden ons bezig met soorten data die worden gevangen in op maat gemaakte datastructuren. Zijn wij 40-plus en hebben wij de tijd van modelgedreven systeemontwikkeling nog meegemaakt dan hebben wij vaak nog wel belangstelling voor metadata. Maar; "dat zijn data-over-data en niet data-over-gebruikersobjecten" en dus héél iets anders – NOT.

Als iemand die zijn carrière heeft gebouwd op (meta)datagedreven systemen weet ik tot in mijn ruggemerk hoe onzinnig het onderscheid tussen data enerzijds en datastructuren anderzijds is. Desondanks blijf ik een database expert en geen gebruiker en juist daarom is werken met content experts (spreek uit: gebruikers) zo boeiend. De meest onvergetelijke ervaring was bij een overheidsorganisatie die destijds de arbeidsbemiddeling deed. Deze club beschikte over honderdduizenden vacatures en CV's die zo goed mogelijk aan elkaar moesten worden gekoppeld. De belangrijkste verbinding tussen vraag en aanbod was het begrip 'beroep'. Men kende in de diverse systemen zo'n 2700 verschillende beroepen, vaak met verschillende ID's en dus met een enorme datapuinhoop als resultaat. Die beroepen werden verzameld en gespecificeerd door een gespecialiseerd bedrijf à raison van een paar miljoen euro per jaar. Aan ons ICT'ers om vraag en aanbod op elkaar te laten aansluiten door het masterdata management op poten te zetten. Behalve dat er duizenden beroepen waren waaraan de honderdduizenden CV's en vacatures waren opgehangen, was het probleem vooral dat er ook een goede match kan zijn wanneer niet precies hetzelfde beroep wordt aangeboden als waarnaar wordt gevraagd. De masterdata leverancier leverde daarom niet alleen beroepen maar ook beroepsverwantschappen. Via een listig mechanisme werd voor elk van de 7.287.300 mogelijke beroepcombinaties een verwantschapscore bepaald en die verwantschap werd dan weer gebruikt om tot een goede match te komen. Dat gebeurde met behulp van fuzzy logic, zodat niemand hoefde te weten hoe het werkte.

De eerste uitdaging was om tot één beroepenlijst te komen. Naast alle technische problemen was er één contentprobleem: wat te doen met het beroep van prostituee? Men wilde dit beroep elimineren. Ik verzette mij daar tegen: het oudste beroep ter wereld is in Nederland legaal en bepaald niet zeldzaam en de overheid heeft ook een taak voor sexwerkers die immers

ook belasting betalen. Besloten werd om het beroep te houden maar alle verwantschappen náár prostituee te elimineren. Dus: als een prostituee haar CV opgeeft dan krijgt zij vacatures te zien voor, zeg, receptioniste of verkoopster. We hadden echter geen rekening gehouden met de ICT-organisatie. Die bleek vanouds programmatuur te hebben gemaakt waarin de verwantschapsrichting was omgewisseld. Argeloze receptionisten en verkoopsters werden opeens geconfronteerd met vacatures van sexhuizen. Tegelijk konden prostituees niet uit het leven stappen, althans niet via het matchingsysteem. Nog vervelender werd het toen een vestigingsmanager het een aardig idee vond om een vacature voor prostituee op te hangen in de etalage. De Christelijke partijen in de Tweede Kamer boos, want het ging om de officiële overheidsmatchmaker en men was bang dat het begrip 'passend werk' zou worden opgerekt.

De 'sexwerkersincidenten' hadden ons en de gebruikers wel aan het denken gezet over de verwantschappen en we besloten om samen eens te gaan kijken naar de kwaliteit ervan. Die bleek zo dramatisch slecht dat de gebruikers zich afvroegen hoe bepaalde werkzoekenden überhaupt zouden kunnen worden gematcht – met of zonder fuzzy logic. Eén query later wisten we dat er inderdaad tienduizenden productiemedewerkers in de database zaten waarvoor geen vacatures uitstonden en die niet werden gekoppeld aan de vele CV's voor beroepen variërend van vorkheftruckchauffeur tot micaspiljer. Je geeft miljoenen per jaar uit aan masterdata maar dan heb je ook niks. En dus zochten en vonden we samen met de beroepkundige gebruikers een andere oplossing: die honderdduizenden CV's bleken een schatmijn aan verwantschappen te bevatten. Met een handvol SQL query's konden we kwalitatief goede verwantschappen boven water krijgen waarmee 80 procent van de werkzoekendenpopulatie kon worden gematcht, natuurlijk met die briljante fuzzy logic engine. Eendrachtige samenwerking tussen ICT'ers en gebruikers had de kwaliteit van de matching enorm verhoogd en de kosten van masterdata acquisitie verlaagd.

Eind goed, al goed? Ik vrees van niet. Wij zijn alweer jaren weg, de organisatie bestaat niet meer, maar het systeem nog wel. Een tijdje terug heb ik mijn CV ingevoerd. Ik ben econoom, ICT-specialist en docent. Kennelijk ben ik een probleemgeval want wekelijks kreeg ik een mailtje dat er helaas geen vacatures voor mij waren. Als ik jong en knap was zou ik bijkans overwegen om maar de prostitutie in te gaan.

Dr. R.J. Veldwijk is partner bij FAA Partners.