

Stage in AI: Het frameprobleem

Robbert Lindeman, CKI student aan de UU
(robbert.lindeman@phil.uu.nl)

het Onttrafelen van

Robbert Lindeman is CKI-student aan de Universiteit van Utrecht. Hij heeft daar een onderzoeksstage gedaan naar het frameprobleem van de AI. “Het wiel is een uiterst efficiënte manier om je voort te bewegen, maar Moeder Natuur heeft geen wielen voortgebracht. Om een artificiële agent te bouwen moet er dus een biologisch plausibel model te vinden zijn waarvan men gebruik kan maken.” Robbert Lindeman heeft samen met mede CKI’er Pim Soomer gewerkt aan een systeem (FrameDB) dat de aangedragen oplossingen voor dit frameprobleem probeert te ordenen.

Na een lezing, waar de term frameprobleem weer eens per ongeluk was gevallen, was ik met een vriendin wat aan het napraten. Het was een leuke lezing, dus in een vlaag van nawerkend enthousiasme kwamen we tot de conclusie dat dat rare frameprobleem toch echt niet zo moeilijk kon zijn. Gewoon een kwestie van wat dwarsverbanden leggen tussen psychologie, filosofie en informatica, een beetje formaliseren en klaar! Het idee om het probleem breed aan te pakken sprak me erg aan en wellicht zat hier een onderwerp voor multidisciplinair onderzoek in. Maar al dwalend over het internet kon ik geen enkel mooi overzicht vinden waarin het frameprobleem overzichtelijk uitgelegd stond. De gegevens op *Wikipedia* en de *Stanford Encyclopedia of Philosophy* waren op dat moment nogal mager en lieten geen van beiden het eigenlijke probleemgebied zien. Hierdoor kreeg ik het gevoel dat er onder de oppervlakte meer zat. Het werd dus tijd om in de literatuur te duiken...

Ontwikkeling van de term

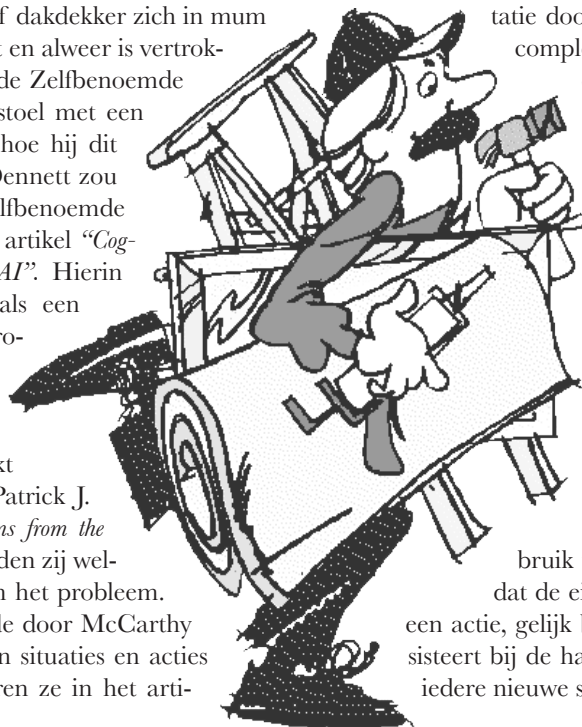
Altijd als men doe-het-zelft wordt het een janboel. Waar de professionele behanger, elektricien of dakdekker zich in mum van tijd van zijn taak gekweten heeft en alweer is vertrokken in zijn witte bedrijfswagen, zit de Zelfbenoemde Kluskoning nog steeds in zijn klapstoel met een mok thermoskoffie te bekostigen hoe hij dit varkentje eens zal wassen. Daniel Dennett zou je kunnen zien als een dergelijke Zelfbenoemde Kluskoning. In 1984 schreef hij het artikel “*Cognitive Wheels: The Frame Problem of AI*”. Hierin herkende hij het frameprobleem als een “nieuw, diep epistemologisch probleem” dat “door generaties van filosofen” onopgemerkt gebleven was; een fikse claim.

Het frameprobleem werd ontdekt in 1969 door John McCarthy en Patrick J. Hayes. In “*Some Philosophical Problems from the Standpoint of Artificial Intelligence*” wijdden zij welgeteld één pagina en drie regels aan het probleem. In het artikel bouwen zij voort op de door McCarthy ontwikkelde *Situation Calculus*, waarin situaties en acties centraal staan. Een situatie definiëren ze in het arti-

kel als een momentopname van de wereld, volledig beschreven in een set logische proposities. Proposities die veranderd kunnen worden door een actie heten *fluents*. Twee opeenvolgende situaties worden gescheiden door een actie. Stel dat onze Kluskoning een spijker die 3 centimeter in de muur zit, 4 centimeter de muur in wil krijgen, dan moet hij hiervoor een actie uitvoeren met zijn hamer, waardoor de spijker 1 centimeter verder de muur in geraakt. Een actie wijzigt dus de waarheidswaarde van een propositie van de ene beschrijvende set naar de volgende.

Een artificiële agent kan met behulp van de Situation Calculus een omgeving representeren en acties plannen. Maar als de situatie iedere keer volledig beschreven moet worden, hoe zit het dan met de objecten in de wereld die geen slachtoffer zijn van de hameractie? De positie van de klapstoel die toevallig aan een tafel in de buurt van de muur staat, zal als gevolg van de actie met de hamer waarschijnlijk net zo min gewijzigd worden als de hoeveelheid koffieaanslag in de mok op die tafel. Er moeten dus condities gegeven worden die ervoor zorgen dat acties irrelevante fluents niet veranderen. Dit omdat de Situation Calculus vereist dat de situatie volledig beschreven is. Stel dat er n mogelijke acties zijn en m fluents, dan moeten er in totaal $m \times n$ van zulke condities zijn. Naarmate een representatie door meer mogelijke acties en voorwerpen complexer wordt, neemt de hoeveelheid condities ook flink toe. Deze condities heten *frame-axioma's*. De complexiteit van het bepalen van die axioma's vormt het frameprobleem.

Het hierboven beschreven probleem komt dus voort uit het gebruik van de Situation Calculus. Voor informatici een irritant engineeringprobleem (een technisch obstakel zoals zovele) en voor de overige mensen binnen de harde AI een fijne puzzel om de tanden in te kunnen zetten. Een oplossing waarmee men op de proppen kwam was het gebruik van zogenaamde *non-monotone logica*. Deze vorm van logica maakt gebruik van persistentieregels die ervoor zorgen dat de eigenschappen die onafhankelijk zijn van een actie, gelijk blijven. De locatie van de klapstoel persisteert bij de hamerslag. Op die manier hoeft je niet in iedere nieuwe situatie alle frame-axioma's na te lopen.



een Gordiaanse Knoop?

Na enige jaren kwam Dennett om de hoek kijken. Hij vroeg zich af of mensen het frameprobleem eigenlijk wel opgelost hadden. Wij zijn per slot van rekening ook een soort agenten in een omgeving. Moeten wij ook frame-axioma's formuleren als we een actie gaan plannen? Kunnen we niet gewoon aannemen dat de rest nog steeds normaal is? Intuïtief is dit juist wat wij doen en daarom is er volgens Dennett helemaal geen sprake van een frameprobleem. Wij hoeven helemaal geen enorme lijst van frame-axioma's te formuleren. Maar wat dan als er iets abnormaals gebeurt? Zo kan het zijn dat tijdens het spijkeren de kop van de hamer afvliegt. Dit is in zoverre abnormaal omdat je er vanuit gaat dat je te allen tijde met een hamer kunt hameren, en je dit tot één tik geleden ook hebt gedaan, terwijl het nu niet meer kan. Mensen kunnen met zo'n defecte hamer omgaan, omdat we blijkbaar tóch bijhouden hoe normaliter de situatie om ons heen eruit ziet. We weten het meteen te herkennen als iets ongewoons plaatsvindt. Het probleem zit hem nu volgens Dennett in het aan een systeem uitleggen van de normale situatie.

Nu zou je kunnen zeggen dat er twee verschillende doelen zijn voor AI. Voor de één zal het doel het maken van een zo intelligent mogelijke machine zijn; maakt niet uit hoe, als het maar intelligent is. De ander zou graag willen dat AI leidt tot inzicht over de mens; als er een machine gemaakt wordt, dan moet deze zeker eenzelfde denkraam als wij hebben. Dennett pleit ervoor geen cognitieve wielen uit te vinden als oplossing voor het frameprobleem. Het wiel is een uiterst efficiënte manier om je voort te bewegen, maar Moeder Natuur heeft geen wielen voortgebracht. Om een artificiële agent te bouwen moet er dus een biologisch plausibel model te vinden zijn waarvan men gebruik kan maken.

Dennett stelt dat een oplossing als non-monotone logica zo'n cognitief wiel is. Het is dus geen juiste methode om onze cognitie te modelleren, hoe elegant de methode er misschien ook uit ziet. Het frameprobleem is een epistemologisch probleem dat breed binnen de cognitiewetenschap aangepakt dient te worden. Om onderscheid te kunnen maken tussen het oude probleem, zoals beschreven door McCarthy & Hayes, en het nieuwe probleem, heeft Dennett het oude probleem omgedoopt tot het *Frame Problem - Proper*. Dennett's *whole pudding* gaat vanaf dat moment onder de term *frameprobleem* verder.

Uiteraard is er nogal wat kritiek gekomen op Dennett voor het vergroten van het probleemgebied. Er was een tamelijk heldere definitie van het oorspronkelijke probleem, dat nu is vervangen door een probleem met een open eind, waarbij niet duidelijk is hoe een eventuele oplossing eruit zou moeten zien (als die al

zou bestaan). Menig wetenschapper zag in Dennett's aanpassingen geen, of juist heel veel heil. De eigen definities vlogen over het papier, de één nog feller verdedigd dan de ander. De meest memorabele verdediging staat op naam van Hayes die in zijn "What the Frame Problem Is and Isn't (1987)" nogal expliciet van wal steekt tegen Jerry Fodor.

Dennett heeft er dus een bende van gemaakt door de keurig geordende gereedschapskist van McCarthy & Hayes (model Frame) om te stoten. Vriend en vijand komen helpen met opruimen en zij zijn genereus met het aanbieden van zakken spijkers, plamuurmessen en lijmpistolen, maar er is geen consensus over wat nu eigenlijk wel en niet in de kist thuis hoort.



Tijd voor een nieuwe gereedschapskist

Ik had het idee een overzicht te maken van het frameprobleem als voorstel voor een onderzoeksstage ingediend bij de faculteit Wijsbegeerte aan de Universiteit Utrecht, waar ik gelukkig aan de slag kon. De bedoeling was om een database met webinterface te maken, waarin de literatuur-

structuur van het frameprobleem gevangen kon worden. Zo zou je bij de gegevens uit Dennett's "*Cognitive Wheels: The Frame Problem of AI*" (naast algemene informatie als auteur, jaartal, etc.) zien welke onderwerpen in het artikel besproken worden, hoe dit artikel reageert op andere artikelen (alsmede het onderwerp waar die reactie mee gepaard gaat), maar ook hoe andere onderzoekers op het artikel gereageerd hebben. Op deze manier wordt niet alleen weergegeven hoe vaak iemand geciteerd is (zoals bij *Citeseer [een zoekmachine gespecialiseerd in het vinden van wetenschappelijke documenten, red.]*), maar ook waarom iemand geciteerd is.

Uiteraard moeten aan het frameprobleem gerelateerde onderwerpen ook opgenomen worden. Van een aantal deelonderwerpen in deze literatuur zijn echter meerdere definities waardoor het niet gelijk duidelijk is hoe termen worden gebruikt. Door termen te objectiveren kunnen die definities tenminste met elkaar vergeleken worden en door de terminologie in te lijven in een webarchitectuur wordt die informatie ook nog eens makkelijker toegankelijk. Kortom we hebben een wiki (*Een open informatiebron naar het model van internetencyclopedie Wikipedia.org, red.*) nodig.

Het frameprobleem: een irritant engineeringprobleem voor informatici en voor de overige mensen binnen de harde AI een fijne puzzel om de tanden in te kunnen zetten.

Ten slotte zou het weergeven van nog openstaande deelproblemen, het onderzoeksgebied van het frameprobleem in kaart brengen. Zo kan iedere cognitiewetenschapper, al dan niet in spe, zien waar er nog werk verricht moet worden. Voor studenten is dat extra interessant, want als er een onderwerp voor een onderzoek nodig is, dan kan hij of zij op de website een voor hem of haar interessant onderwerp kiezen uit een lijst met nog openstaande problemen. Tijdens het doen van het onderzoek brengt de student mis-

het licht die vervolgonderzoek rechtvaardigen, of het huidige onderzoek besluiten. Deze resultaten kunnen dan ingevoerd worden, waardoor het verslag of essay wordt opgenomen in de literatuurstructuur. De wiki wordt aangepast voor zover er informatie winst is en ideeën voor vervolgonderzoek worden toegevoegd. Het resultaat is een soort doorgeefonderzoek onder studenten waarbij anderen de resultaten van de vorige onderzoeken toepassen. De opgedane kennis verdwijnt

dus niet langer in een stoffige kast!

Zoals je ziet lag er een flinke klus voor de boeg. Gelukkig vond ik in mijn mede-CKI'er Pim Soomer een fijn brein om mee te werken aan dit project. Ik zou me bezig gaan houden met de literatuurstructuur en de probleemstellingen binnen het frameprobleem, en hij zou zich richten op het bouwen van de wiki. We kregen de beschikking over een server, waarop we de database zouden kunnen gaan inrichten. Dat was eigenlijk best spannend. Allebei wisten we nog niets over databases en serverscripting, wat op zich een hoop werk was om uit te zoeken, maar daardoor hadden we wel het gevoel dat we onbevangen aan de slag waren. Het maken van het systeem zou uiteindelijk resulteren in ongeveer 4 maanden programmeerwerk. Dat klinkt wellicht wat droog, maar het is heel erg fijn om iets te zien groeien dat je zelf aan het maken bent. Toen we ongeveer halverwege waren besloten we ons kindje FrameDB te noemen.

Maar dan is er nog het probleem van het vullen van het FrameDB. Het idee klinkt aardig, maar al die informatie komt er niet vanzelf in! Als welkome afwisseling op al het programmeren hadden we met onze tweede begeleider afgesproken om eens per week artikelen over het frameprobleem te bespreken. Hierin zouden we kijken welke informatie we uit artikelen kon-

den halen om in ons systeem te zetten. Van deze besprekingen hebben we heel veel geleerd, van zowel het frameprobleem zelf, alsook hoe je vermeende oplossingen herkent als verplaatsingen van het probleem naar een ander vakgebied. Daarnaast verloren we door de wekelijkse besprekingen niet uit het oog waar we het allemaal voor deden.

Soms kon ik me wel eens blind staren op al dat programmeerwerk. Uiteindelijk hebben we informatie uit ongeveer 20 artikelen

opgenomen in het systeem.

Helaas is het deel van FrameDB met de probleemstellingen wegens tijdgebrek een beetje een ondergeschoven kindje geworden, maar op zich is het goed te gebruiken. Daarbij is het raamwerk van FrameDB niet specifiek geschikt om alleen de literatuurstructuur voor het frameprobleem aan te kunnen, het kan zelfs relatief gemakkelijk uitgebreid worden om literatuurinformatie uit allerhande onderzoeksgebieden te bevatten! Hoewel we nog genoeg ideeën hebben om FrameDB uit te breiden, is het helaas tijd om te gaan afstuderen. Hopelijk heb ik de AI'ers van Nederland ervan kunnen overtuigen dat het frameprobleem een mooi probleem is om je tanden in te zetten. En mocht je door dit stukje zelfs zin hebt om verder te gaan met FrameDB: we hebben een mooie handleiding van 32 pagina's voor je klaar liggen... ☺

Bekijk het resultaat op fermat.phil.uu.nl (Beta)

Literatuur

- Dennett, D.C. "Cognitive Wheels: The Frame Problem of AI." In *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, edited by Pylyshyn, Z.W., 42-65. Norwood, NJ: Ablex, 1987.
- Hayes, P.J. "What the Frame Problem Is and Isn't." In *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, edited by Pylyshyn, Z.W., 123-137. Norwood, NJ: Ablex, 1987.
- Fodor, J.A. "Modules, Frames, Fridgeons, Sleeping Dogs, and the Music of the Spheres." In *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, edited by Pylyshyn, Z.W., 139-149. Norwood, NJ: Ablex, 1987.
- McCarthy J, Hayes P (1969) *Some philosophical problems from the standpoint of artificial intelligence*. *Machine Intelligence* 4:463-502 (niet zuiver Chicago Citation Style)