

Ontologie, EEN MANIER om kennis TE structureren

Huub Scholten

Leerstoelgroep Informatie Technologie Wageningen UR, 0317 – 48 46 31, huub.scholten@wur.nl

Samenvatting

Vijftien jaar geleden had vrijwel niemand van het begrip ontologie gehoord en was het een begrip waar alleen filosofen wat mee konden. Nu is het een buitengewoon populair onderwerp op internet. Dat het iets met kennistechnologie te maken heeft, en met het intelligent opslaan van data, weten sommige mensen nog wel. Zijn ontologieën zo iets als databases? Moet ik hier iets van weten? Heb ik er iets aan? Op deze vragen geeft Huub Scholten in dit artikel antwoord.

Inleiding

Over het nut van computers is iedereen het tegenwoordig wel eens. Anders wordt het wanneer computers iets met kennis moeten doen. In de jaren 50 werden de eerste kennissystemen ontwikkeld, maar een groot succes was dat niet. Alleen duidelijk definieerbare, samenhangende en goed omschreven stukjes kennis werden op deze manier vastgelegd en ter beschikking van gebruikers gesteld. Nu, 50 jaar later, is de situatie anders. Internet speelt een centrale rol in de uitwisseling van informatie en kennis, maar deze informatie is heel divers en erg ongestructureerd, waardoor het moeilijk is deze op een doelmatige manier te gebruiken. Als u nagaat hoe snel zoekmachines als Google uw zoekopdracht succesvol uitvoeren, dan zult u zich misschien wel eens hebben afgevraagd hoe er op miljarden pagina's WWW zo snel de gezochte informatie gevonden wordt. De wonderolie om dat mogelijk te maken heet "ontologie" en de meest korte beschrijving ervan zou als volgt kunnen luiden: een manier om kennis te structureren en op te slaan die voor mensen en computers 'begrijpelijk' is.

Twintig jaar geleden betekende het woord ontologie alleen nog maar 'esoterisch onderdeel van de filosofie, de zijnsleer, over wat er in de wereld 'is'. De term is ontleend aan Aristoteles. Die onderscheidde theoretische filosofie, bestaande uit fysica, ontologie en logica en daarnaast de praktische filosofie bestaande uit ethica, politiek en poëtica. De ontologie bestudeert de *zijnden* als zodanig, hun aard, eigenschappen, en onderlinge betrekkingen. Ontologie heeft sinds ongeveer 1990 een heel andere betekenis gekregen. In mei 2002 vond de zoekmachine Google 300.000 webpagina's op het trefwoord 'ontologie', in mei 2003 waren dat 700.000 webpagina's. De term is overgenomen door kennisingenieurs en heeft alleen nog vagelijk te maken met wat Aristoteles er onder verstond. De meest gehanteerde definitie

is van Gruber (1993, 1994) en luidt (vertaald) als volgt: *een ontologie is een expliciete specificatie van een conceptualisering, waarbij concepten dingen zijn die kunnen worden weergegeven of waarover kan worden gepraat*. Borst (1997) voegt hier nog aan toe dat over deze concepten en de relaties ertussen consensus moet bestaan. Dit levert de volgende definitie op: *een ontologie is een formele specificatie van een gemeenschappelijk conceptualisering*.

In dit artikel hoeft u geen samenvatting te verwachten van 700.000 pagina's internet. Hooguit een korte inleiding over het *wat* en het *waarom*. Verder ga ik in op mijn persoonlijke betrokkenheid bij ontologieën.

Waarom een ontologie ontwikkelen?

Sinds de begintijd, zo'n 15 jaar geleden, zijn ontologieën in ras tempo vanuit het *Artificial Intelligence Lab* terechtgekomen op het bureau van de domeindeskundige, zoals u en ik. Op WWW kom je tegenwoordig veel ontologieën tegen. Vaak zijn dat simpele, maar grote taxonomieën, die bijvoorbeeld websites categoriseren zoals die van Yahoo. Het WWW Consortium (W3C) is met een groot project bezig om een ontologische beschrijvingstaal te definiëren voor het beschrijven van kennis op webpagina's zodat 'software agents' op het web deze kennis kunnen 'begrijpen'. Ook zijn er veel militaire toepassingen, genetische toepassingen, ontologieën voor de zakenwereld (Uschold *et al.*, 1998), terminologische ontologieën voor producten en diensten en de protocollaire geneeskunde (de Clercq *et al.*, 2001). Het beschrijven van planning in de voedselketen (Houba *et al.*, 2000) zou niet kunnen bestaan zonder een stevige ontologische basis.

Essentieel bij een ontologie is dat mensen er wat mee kunnen maar machines (computers) ook. En je hoeft geen AI (*Artificial Intelligence, géén Agro Informatica*) *nerd* te zijn om er wat aan te hebben. Een ontologie structureert kennis op een begrijpelijke, maar formele manier. Samengevat zou je kunnen zeggen dat ontologieën gemaakt worden om de volgende redenen:

- *Het realiseren van begrip tussen mens en computer (met name web agents) over informatie structuren*. Als twee e-commerce websites gebaseerd zijn op dezelfde ontologie, snappen *software agents* de termen die gebruikt worden en kunnen ze informatie verzamelen en aggregeren afkomstig van beide websites. Ze begrijpen niet echt wat er staat, maar de informatie is voldoende gestructureerd en bevat voor

de agent hanteerbare semantiek (betekenis) die overigens ook voor de mens begrijpelijk is.

- *Hergebruik mogelijk maken van domeinkennis.* Dit is een van de belangrijkste drijfveren achter het ontwikkelen van ontologieën. Een voorbeeld is het begrip *tijd* dat in veel modellen een rol speelt. Hier spelen allerlei deelbegrippen een rol, zoals tijdsinterval, punten in de tijd, eenheden (seconde, minuut, uur, dag, maand, jaar). Als iemand een (goede) ontologie voor *tijd* heeft gemaakt, kunnen andere deze voor hun eigen ontologie zo hergebruiken.
- *Het expliciet maken van aannames.* Als in een ontologie de aannames gestructureerd zijn opgeslagen, is het makkelijk om deze aan te passen. In de applicatie die gebruik maakt van de ontologie hoeven dan geen hard gecodeerde aannames te worden aangepast.
- *Het scheiden van domeinkennis en operationele kennis.* Als het productieproces voor bijvoorbeeld tractoren is gedefinieerd in een ontologie, onafhankelijk van het product zelf, kan de ontologie ook gebruikt worden voor het produceren van andere producten zoals bijvoorbeeld PC's.
- *Het analyseren van domeinkennis.* Met behulp van een specificatie van domeinkennis in een ontologie, kan gemakkelijk het consequent gebruik van termen worden gecontroleerd.

Meestal is het maken van een ontologie een doel op zichzelf. Vaak is er sprake van een proces dat ongeveer als volgt in elkaar zit:

1. Er wordt een ontologie gemaakt van een stuk domeinkennis. Dat vormt de structuur van de uiteindelijke kennisbank.
2. Met behulp van een op de ontologie gebaseerde *tool* wordt concrete domeinkennis verzameld.
3. Deze kennis wordt opgeslagen als *instantie* van de ontologie in een kennisbank (*knowledge base*).
4. Er wordt een applicatie gemaakt die deze kennis gebruikt (die dus samenwerkt met de kennis in de kennisbank).

Voor ik in zal gaan op de voordelen van deze aanpak, door deze te vergelijken met een minder flexibele aanpak waarbij de kennis is vastgelegd in een database, moet ik eerst wat meer vertellen over wat ontologieën zijn. Een goede inleiding in ontologieën is te vinden in Chandrasekaran *et al.* (1999)

Hoe zit een ontologie in elkaar

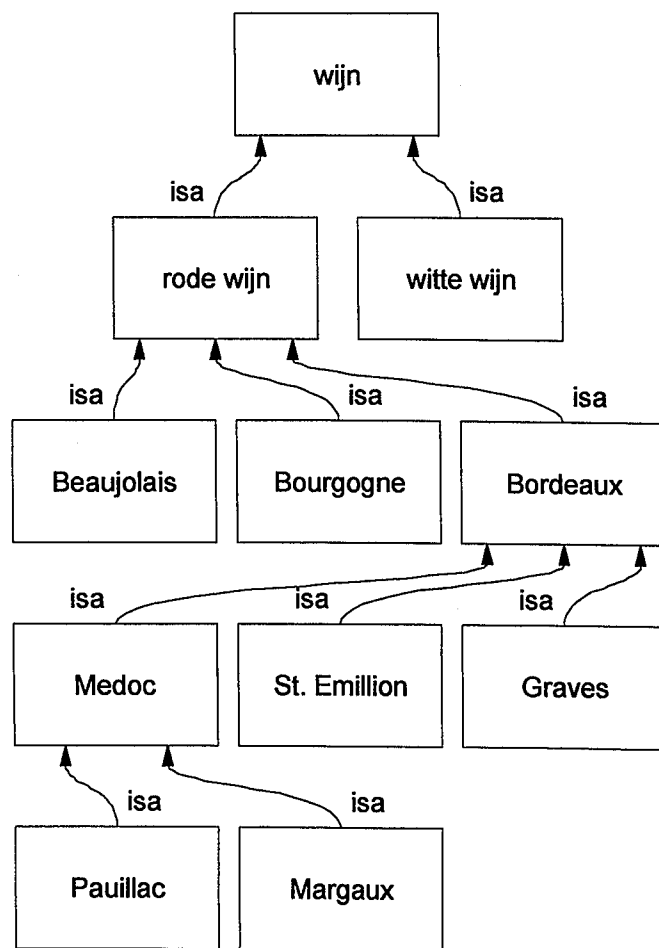
Een aantal begrippen zijn essentieel voor een ontologie, maar de manier waarop deze worden besproken is vaak afhankelijk van de gebruikte software om de ontologie te maken of er mee te werken. Ik zal proberen zo algemeen mogelijk te zijn en alleen als het nodig is afwijkende woorden te gebruiken, die ontleend zijn aan objectoriëntatie, al betekenen ze daar niet precies hetzelfde als in de wereld van de ontologie.

Om de semantiek (betekenis) van data te kunnen beschrijven, moeten ontologieën beschikken over een standaard vocabulaire, dat bepaalt wat er in de kennisbank over een bepaald domein uiteindelijk kan worden uitgedrukt. Centraal in dat vocabulaire staat het begrip 'concept', dat staat

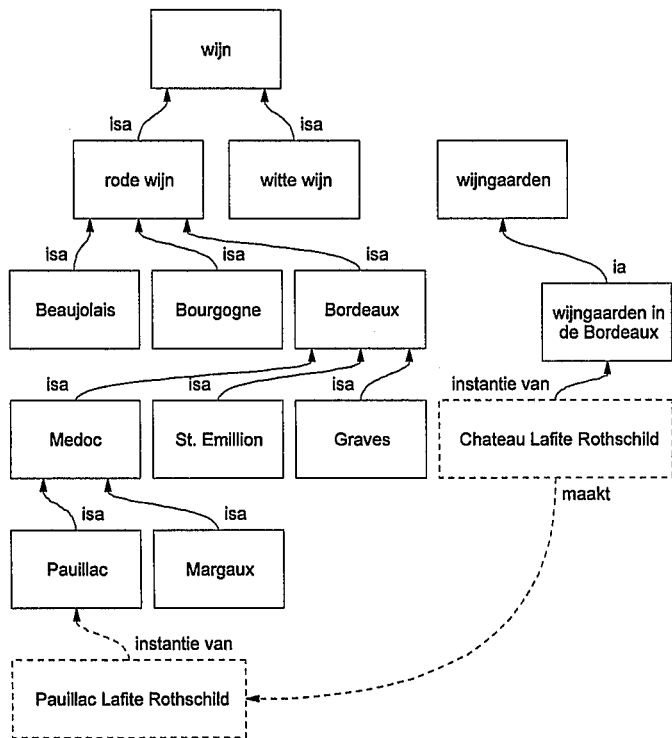
voor dingen waarover gediscussieerd wordt en die moeten worden gerepresenteerd. Het begrip 'concept' is dus veel ruimer dan bijvoorbeeld 'entiteit' en het omvat abstracte en concrete dingen, maar ook processen en taken. Concepten worden gebruikt om deze termen te definiëren en uit te leggen. Een tweede wezenlijk begrip is 'relatie'. Relaties ordenen de concepten in een structuur. Dit kan bijvoorbeeld een hiërarchie zijn. Een ontologie kan worden gezien als een raamwerk dat de semantiek (betekenis) weergeeft van data/kennis over een bepaald domein op een manier waarmee zowel mensen (niet alleen AI-technologen, maar ook domeindeskundigen) als computers uit de voeten kunnen. Naast concepten wordt de structuur van de data/kennis bepaald door relaties tussen de concepten. De relaties kunnen tot allerlei (ook zelf te definiëren) typen behoren. Een heel simpele hiërarchie kan er als volgt uitzien als in figuur 1.

Naast 'concepten' en 'relaties' zijn er vaak ook nog andere elementen. Vrijwel altijd is er sprake van 'instanties'. Een instantie is een klasse die een specifieke invulling vormt van een categorie: een individuele representant van een soort (figuur 2).

Voor ontologieën heb je niet alleen termen nodig, maar ook een *format* waarin een ontologie wordt opgeslagen en bij voorkeur ook een *tool* om de ontologie te maken. Er zijn vele *formats* waar ik hier niet op in zal gaan, want als je een goede *tool* gebruikt, is het niet al te moeilijk om van het ene *for-*



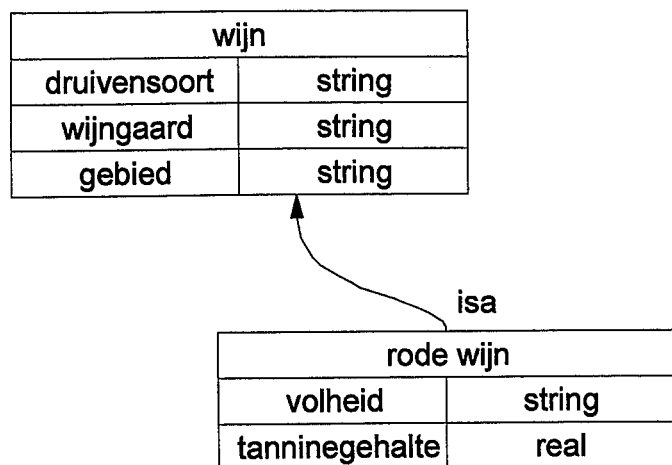
Figuur 1. Een deel van een hiërarchische ontologie van wijn. Een *isa*-relatie betekent 'is-een'.



Figuur 2. De classes van de ontologie van Figuur 1 uitgebreid met een paar instanties (gestippeld) en andere relaties dan 'isa'.

mat naar het andere te gaan (al verlies je dan soms wel informatie). De meest geavanceerde tool op dit moment is Protégé2000, maar er wordt hard gewerkt aan andere nog betere tools. Zelf heb ik voor Protégé2000 gekozen vanwege het gebruiksgemak, de stabiliteit (er wordt al ruim 10 jaar aan gewerkt), het open karakter (open source en de source code is ook beschikbaar), waarbij het eenvoudig is om je eigen uitbreiding te maken, de platform-onafhankelijkheid van de tool (Java) en de prettige gebruikersgroep van deskundigen die elkaar helpen.

De keuze van een format en een tool heeft ook consequenties. Tools gebruiken andere termen voor de begrippen concept en relatie. Zo heten concepten in Protégé2000 classes en het begrip slot wordt gebruikt om relaties mee te definiëren, maar ook om eigenschappen (properties) van classes (=concept



Figuur 3. Rode wijn 'erft' van wijn ook de eigenschappen 'druivensoort', 'wijngaard' en 'gebied'.

ten) te specificeren. Classes kunnen subclasses hebben die de eigenschappen erven van de parent class. Een voorbeeld: als een class het begrip 'wijn' beschrijft, kunnen twee subclasses witte wijn en rode wijn beschrijven. De relaties tussen enerzijds witte wijn en rode wijn met de class wijn is een slot van type isa (wat staat voor 'is een'). Basale slots als isa hoeven niet gedefinieerd te worden, want ze zijn al beschikbaar, gewoon omdat Protégé2000 erg gebaseerd is op begrippen die ontleend zijn aan objectoriëntatie, inclusief vererven. Eigenschappen (properties) van concepten (=classes) worden in Protégé2000 ook gemodelleerd met behulp van slots (figuur 3).

Zodra er met een goede tool een ontologie gemaakt is, kan deze voor allerlei toepassingen worden gebruikt. Bijvoorbeeld als datastructuur of om een kennisbank te maken door er een instantie van te maken die vervolgens gevuld wordt met kennis. De gevulde kennisbank wordt meestal opgeslagen in een intern format, maar kan heel gemakkelijk worden geëxporteerd naar bijvoorbeeld XML.

Database of ontologie

Omdat iedereen wel iets van databases en datamodellen weet of denkt te weten, zal ik hier ontologieën en databases met elkaar op de volgende punten vergelijken:

- Expliciteit
- Flexibiliteit
- Syntactische en semantische integriteit
- Representatief vermogen

In een database kan een veld plaats voor van alles en nog wat worden gebruikt: plaats van een adres, plaats waar een politieauto of een boot zich op een bepaald moment bevindt. In een ontologie wordt onderscheid gemaakt tussen elk van deze begrippen plaats, op basis van een algemene (daarom liggende) beschrijving van het gemeenschappelijke in het begrip plaats. Op deze manier maakt een ontologie het data-model meer expliciet.

Met behulp van ontologieën is het mogelijk (zeer) verwante domeinen te beschrijven, wat met behulp van datamodellen (E/R-diagrammen, UML) heel moeilijk, zo niet onmogelijk is. Een ontologische aanpak is ook in andere opzichten meer flexibel. Op basis van nieuwe inzichten is het vrij gemakkelijk om een ontologie aan te passen. De daarop gebaseerde kennisbank/database is dan automatisch aangepast en correct.

Het gebruik van SQL en ODBC waarborgt een syntactische integriteit in databases. Om semantische integriteit te bereiken worden allerlei afspraken gemaakt, zoals DES (Data Element Standard). Dit is een verzameling afspraken binnen een organisatie over te gebruiken termen en hun betekenis. Een ontologie kan worden gebruikt als formele DES, die ook nog eens kan worden getest op volledigheid en consistentie. Op deze manier maakt een ontologische aanpak ook semantische integriteit mogelijk. Een traditionele DES kan zo iets niet. Metadata repositories zijn een variant op een DES, maar ook deze hulpmiddelen kunnen niet wat nodig is, omdat de metadata niet veel meer zijn dan commentaar op het data-model. In ieder geval zijn ze verder beperkt tot het data-model waar ze voor ontwikkeld zijn, terwijl een ontologie juist gemaakt wordt voor hergebruik.

In tegenstelling tot databases/datamodellen kunnen ontologieën niet alleen een taxonomie beschrijven van entiteiten maar ook van relaties. Voor entiteiten is dat nauwelijks iets bijzonders: een programmeur is een mens, een mens is een zoogdier en een zoogdier is een dier. Een ontologie kan ook dit bevatten: als twee mensen vrienden zijn, dan zijn die twee mensen ook bekenden. Naast dit voorbeeld hebben ontologieën nog veel mogelijkheden om relaties te beschrijven, waarbij de gebruiker zelf het karakter van de relatie kan bepalen. Hierdoor (en door allerlei andere vermogens) is een ontologie veel rijker in representatievermogen dan databases.

Mijn betrokkenheid bij ontologieën

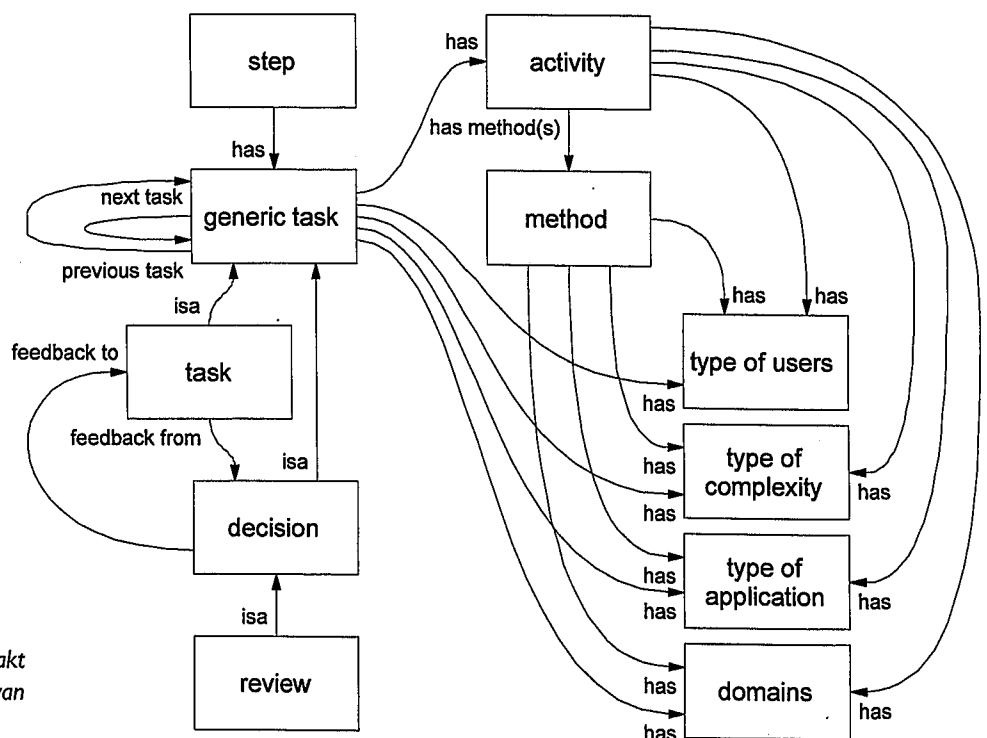
Al zo'n kleine tien jaar ben ik geïnteresseerd in het opslaan van kennis over processen en met name over het maken en werken met simulatiemodellen. Al snel kwam ik op het idee om de kennis over dit proces te structureren met behulp van een ontologie. Eerst is de benodigde kennis verzameld en vastgelegd in een enigszins gestructureerd, maar niet op een ontologie gebaseerd Handboek Good Modelling Practice (Van Waveren *et al.*, 1999, Scholten *et al.*, 2000, 2001). De kennis daarin opgeslagen is weliswaar niet gebaseerd op een formele ontologie, maar wel op een soort informele. De kennis hierin is gestructureerd en het resultaat van discussie en overeenstemming met vertegenwoordigers van alle *stakeholders* in het Nederlandse waterbeheer, waar veel gebruikt gemaakt wordt van modellen. Inmiddels wordt de laatste hand gelegd aan een op dit handboek gebaseerde serie NEN-normen die hetzelfde doel dienen als het handboek, maar die wat formeler zijn geformuleerd.

Terwijl deze initiatieven nog volop in ontwikkeling waren, was het duidelijk dat het handboek noch de normen alle betrokkenen bij het modelleren proces optimaal behulpzaam zouden kunnen zijn. Betrokkenen hebben verschillende rol-

len en vanuit die rollen moeten ze verschillende dingen doen binnen het proces. Ook is niet iedere modelleerklus even ingewikkeld en is het doel van de klus vaak anders, hetgeen invloed heeft op wat er binnen zo'n modelleerklus door de diverse betrokkenen moet worden gedaan. Kortom, genoeg reden om in een nieuw project op basis van de ervaringen opgedaan met het handboek en de normen te proberen alle betrokkenen in het modelleerproces beter te ondersteunen. In januari 2002 is het project HarmoniQuA gestart met als doel het omgaan met simulatiemodellen in het waterbeheer in Europa te harmoniseren. In dit project nemen twaalf organisaties uit tien landen deel met Wageningen Universiteit als coördinator.

Met behulp van de projectdeelnemers, allen ervaren modelleers in het waterbeheer, is een ontologie opgesteld van de kennis van het modelleerproces. Een samenvattende schets hiervan is te zien in figuur 4. In een tweede stap is een deel van de kennis opgeslagen in instanties van de ontologie (invulling van de takenstructuur) om zo een kennisbank te verkrijgen over hoe er volgens de deskundigen moet worden gemodelleerd. In de derde stap die nu ongeveer afgerond is, zijn de instanties verder ingevuld door 'domeindeskundigen' (grondwater, hydrodynamica, waterkwaliteit, etc.). De ontologie is gemaakt met behulp van Protégé2000. De kennis over modelleren is door de domeindeskundigen in de kennisbank gestopt met behulp van 500 *spreadsheets*, maar inmiddels is er een *webbased editor* gemaakt die op een veel eenvoudiger manier dan rechtstreeks met Protégé2000 domeindeskundigen in staat stelt om kennis aan de kennisbank toe te voegen of deze te verbeteren.

Parallel met de tweede en derde stap zijn enkele tools ontwikkeld die de kennisbank op een heel intuïtieve manier aanbieden aan gebruikers. Ook is een tool gemaakt die vastlegt wat de gebruikers doen tijdens het modelleerproces en dit opslaat in een model-journaal met een structuur op basis



Figuur 4. Een deel van de ontologie gemaakt voor HarmoniQuA. Alle eigenschappen van de classes zijn weggelaten.

van dezelfde ontologie als de kennisbank. Uiteraard is er ook andere basisfunctionaliteit beschikbaar (printen van de richtlijnen hoe te modelleren en wat er in een bepaalde modelleerklus feitelijk is gedaan). In de toekomst zal er nog functionaliteit worden ontwikkeld om te leren van eerdere (vastgelegde) modelleerklussen.

Hoewel het project HarmoniQuA nog ruim twee jaar duurt, kan ik wel een voorlopige conclusie trekken over de rol van ontologieën in dit project. In april van dit jaar werden de tools en de kennisbank voor het eerst in Athene ingezet op een seminar voor toekomstige gebruikers. De commentaren op de tools van ruim 25 mensen die hieraan deelnamen konden in een dag worden verwerkt tot grote verbazing van alle aanwezigen. Dit was voor een groot deel te danken aan de flexibiliteit van ontologieën bij het structureren van informatie.

Conclusie

Ik wil hier uiteraard niet beweren dat ontologieën alle problemen met databases kunnen oplossen, maar het zou me niet verbazen als in de toekomst het gebruik van een ontologie in plaats van een datamodel het relationele database-paradigma op zijn kop zal zetten. Representatievermogen, flexibiliteit en hergebruikmogelijkheden zullen de argumenten zijn van deze *paradigm shift*.

Gebruikte literatuur

- Borst, W. N. (1997). Construction of engineering ontologies for knowledge sharing and reuse. University of Twente, The Netherlands, Enschede.
- Chandrasekaran, B., Josephson, J., Benjamins, V. (1999). What are

ontologies and why do we need them? IEEE Intelligent Systems 14, 20-26.

- de Clercq, P. A., A. Hasman, J. A. Blom, and H. H. M. Korsten (2001). The application of ontologies and problem-solving methods for the development of shareable guidelines. Artificial Intelligence in Medicine 22:1-22.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition 5((2)): 199-220.
- Gruber, T. R. (1995). Towards principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies 43:907-928.
- Houba, I.H.G., R.J.M. Hartog, J.L. Top., A.J.M. Beulens and L.N. Berkel (2000). Using recipe classes for supporting detailed planning in food industry: A case Study. European Journal of Operational Research, 122, 367-373.
- Scholten, H., R. H. Van Waveren, S. Groot, F. Van Geer, H. Wösten, R. D. Koeze, and J. J. Noort (2000). Good Modelling Practice in water management. in Proceedings HydroInformatics2000. International Association for Hydraulic Research, Cedar Rapids, Iowa, USA.
- Scholten, H., R. H. Van Waveren, S. Groot, F. Van Geer, J. H. M. Wösten, R. D. Koeze, and J. J. Noort (2001). Improving the quality of model-based decision support: Good Modelling Practice in water management. Pages 223-230 in A. Schumann, J. Xia, M. Marino, and D. Rosbjerg, editors. Regional Management of Water Resources. Proceedings of Symposium S2 of the 6th Scientific Assembly of the International Association of Hydrological Sciences, Maastricht, The Netherlands, July 18-27, 2001. IAHS, Maastricht, The Netherlands.
- Uschold, M, King, M, Moralee, S. and Zorgios, Y. (1998) The Enterprise Ontology The Knowledge Engineering Review , Vol. 13, Special Issue on Putting Ontologies to Use (eds. Mike Uschold and Austin Tate).
- Van Waveren, R.H., Groot, S., Scholten, H., Van Geer, F., Wösten, H., Koeze, R., Noort, J. (1999). Vloeiend modelleren in het waterbeheer. Handboek Good Modelling Practice. STOWA/RWS-RI-ZA, Utrecht/Lelystad, 149 pp.