

Web van betekenis

Door Edith van Gameren, fotografie Getty Images

Het semantisch web. Een term die menig hart sneller laat kloppen, niet alleen van informatici, maar ook van filosofen. De discussie over 'het nieuwe internet' neemt soms religieuze vormen aan waarbij sprake is van 'gelovigen' en 'ongelovigen' in de beloften van web 3.0. Op deze plek: een pragmatisch verhaal over toepasbare technologie.

Het beeld van een cultus, daar wil ik graag van af", zegt VU-hoogleraar Frank van Harmelen. "Semantiek wil niets meer zeggen dan betekenis", legt Van Harmelen uit. "De betekenis van een pagina op het huidige web is voor de computer niet toegankelijk. Zoeken naar 'Harmelen' levert net zo goed mijn persoon op als het gelijknamige dorpje. Het zou al schelen in het zoekwerk als de zoekmachine weet dat er zoiets bestaat als een 'dorp' en een 'persoon'. Een persoon heeft een adres, een werkplek, een telefoonnummer. Dat heeft een dorp niet. Dan heb je nog lang niet uitgelegd wat het echt betekent om een dorp of een persoon te zijn, maar het helpt wél bij het zoeken. En het is een verbetering die met de huidige technologie te realiseren is."

Die overgang naar een nieuwe, in zekere zin 'intelligentere' manier van informatievoorziening via internet voltrekt zich in een geleidelijk proces. Volgens Van Harmelen is de voornaamste uitdaging niet eens technisch, maar sociaal. Hoe worden we het bijvoorbeeld eens over standaarden? Zoals HTML een waterdichte afspraak is die iedere browser begrijpt, moeten er ook voor het semantisch web afspraken zijn voordat er überhaupt van betekenisvol zoeken sprake kan zijn.

Standaardisatie

"W3C (World Wide Web Consortium red.) heeft vastgesteld dat het belangrijk is dat

het web de stap naar semantiek maakt en dat standaardisatie nodig is", zegt Van Harmelen. Onder de W3C-paraplu zijn de talen RDF en OWL ontstaan. Logische, gestandaardiseerde beschrijvingen die als metadata aan informatie worden toegevoegd. Aan de laatste hebben Van Harmelen en zijn collega Guus Schreiber bijgedragen. "OWL en RDF zijn nog een soort babytaal. Je kunt in RDF of OWL net genoeg zeggen

Frank van Harmelen

"De voornaamste uitdaging van het semantisch web is niet technisch, maar sociaal"

om zoekmachines beter te laten werken of personalisatie te kunnen toepassen", aldus Van Harmelen. Inmiddels werkt de technologie in afgebakende omgevingen. Bijvoorbeeld in intranetten van ondernemingen die hun eigen kennis weer de baas willen worden. In een programma van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek genaamd CATCH (Continuous acces to our cultural heritage) wordt semantische technologie gebruikt om de ontsluiting van ons cultureel erfgoed te verbeteren. Het gaat om honderdduizenden zeer diverse objecten. Volgens Van Harmelen is het daarbij

belangrijk om aan te sluiten bij werk dat al gedaan is. "Door een stukje software te ontwikkelen kan bijvoorbeeld een catalogus van de Koninklijke Bibliotheek omgezet worden naar OWL. Met middeleeuwse manuscripten is het gelukt die van de KB en van haar Franse tegenhanger beide om te zetten naar een semantisch-webformaat, zodat je ze met een browser samen kunt doorzoeken. Zoek je in het Nederlands naar werken met vogelmotieven, dan krijg je ook Franse treffers."

Expliciet

Rogier Brussee, senior medewerker bij het Telematica Instituut, houdt zich binnen het MultimediaN ViTa-project en in het bijzonder binnen CiC (Cultuur in Context) bezig met semantisch-webtechnologie. Binnen CiC wordt een ontologie (een schema dat een verband legt tussen verschillende begrippen) van gemakkelikheden zoals film, theater en kermis ontwikkeld. "Er zijn veel invullingen van het begrip semantisch web", zegt Brussee. "Het woord semantiek werkt als een psychologische trigger: eindelijk betekenis! De invulling waarin ik geloof is: maak expliciet wat je wilt zeggen, zodat de computer er iets mee kan en zodat we een goede kans hebben dat we het met elkaar eens worden. Met ontologiën maak je dingen expliciet, je maakt verschillende klassen en deelklassen en je legt eigenschappen vast. Het doel is simpel: interoperabiliteit. Dat is ook het doel van één van de belangrijkste toepassingen van

Is Wikipedia

De Nederlandse Wikipedia telt ruim 320.000 artikelen. De encyclopedie is een van de grote internetsuccessen. Toch pleit Larry Sanger, een van de oprichters, ervoor om de 'echte' deskundigen weer aan de knoppen te zetten. Wikipedia zou op de anti-elitaire manier geen betrouwbaarheid kunnen waarborgen.

Wij vroegen vier deskundigen of de 'volksencyclopedie' nog wel te vertrouwen is.

dit moment, het maken van financiële rapportages met XBRL (eXtensible Business Reporting Language). Dat valt buiten OWL en RDF, maar het is ook een voorbeeld van het streven naar interoperabiliteit en automatische verwerkbaarheid met de hulp van een ontologie.”

Ook Brussee vindt de cultus rond het semantisch web overdreven. “Descriptie-logica's, ontologiën, het klinkt vaak wat diepzinniger dan het eigenlijk is. Wat je bijvoorbeeld wilt zeggen is dat de uitgave van een boek eenduidig vastligt door zijn ISBN-nummer. Door de wereld zo logisch te modelleren, maak je je aannames expliciet, maar het blijft een gesimplificeerd model van de werkelijkheid dat niet beter is dan je aannames.”

Ondanks zijn pragmatische en op sommige punten sceptische houding, ziet Brussee semantiek in het web zeker als een stap vooruit. “Het huidige internet functioneert vooral door reversed engineering: als je

Rogier Brussee

“Ik geloof meer in de computer heel veel ‘domme’ dingen te laten doen. Je laat een graafmachine ook geen archeologisch onderzoek doen”

googelt, stel je jezelf steeds de vraag welke woorden er zouden kunnen staan op een pagina die je kan brengen naar de informatie die je eigenlijk zoekt. Die omweg maken we omdat Google er fantastisch goed in is om character strings op het web te vinden, maar de zoekresultaten moeten we allemaal nog zelf doorlezen. Wanneer het resultaat alleen ‘voor menselijke consumptie’ is, is dat al lastig. Computers kunnen er al helemaal niet mee overweg.”

Data-integratie

Betekenis toevoegen aan het web heeft grote voordelen. De talrijke databases in bedrijven en grote organisaties kunnen eindelijk aan elkaar geknoopt worden en in één beweging, nauwkeuriger doorzocht.

“Bedrijven weten zelf niet meer waar hun eigen expertise zit”, zegt Van Harmelen. “Ze komen om in informatie en weten niet meer welk onderzoek gaande is of al gedaan is.” In bedrijven en andere afgebakende omgevingen wordt de semantische technologie dus al toegepast. De stap naar het voor iedereen toegankelijke web wordt echter ook al voorzichtig gemaakt. “Het Nederlandse Ilse experimenteert met specialistische zoekmachines. Bijvoorbeeld voor auto's. Als jij een Opel Astra zoekt, komt de zoekmachine ook met resultaten die daarop lijken. Als gebruiker merk je aan de buitenkant niets van die semantische technologie, je denkt hooguit ‘goed gevonden’.”

Het verbinden van databases is ook volgens Brussee een grote uitdaging. “In kleine databases is de logica, de ‘zin’, nog vrij gemakkelijk te ontdekken. In grote databases zit die zingeving in steeds grotere applicaties, waardoor we het zicht erop kwijtraken. Er zijn ongelooflijk veel data in ontelbare databases opgeslagen. De hamvraag is: wat kun je daarmee?” Brussee haalt als voorbeeld het elektronisch patiëntendossier aan. Om een doktersrekening te sturen is het voldoende om een naam, geboortedatum, burgerservicenummer en adres in een database vast te leggen als ‘persoon’. Moet echter ook alle behandelinformatie in het dossier worden opgenomen, dan moet er een andere database met een ander soort data – veel tekst – aan worden vastgeknoopt. Ook de betekenis van die tekst moet expliciet gemaakt worden.

Graafmachine

De manier waarop semantische technologie moet worden toegepast, is volgens Brussee niet ‘de computer intelligente dingen laten doen’. “Alhoewel dat natuurlijk wel heel leuk is, geloof ik meer in de computer heel veel ‘domme’ dingen te laten doen. Als je astronomische hoeveelheden data hebt, kun je er alleen op een voor een mens betekenisvolle manier mee omgaan door geautomatiseerde stappen toe te passen. De winst zit in het combineren van simpele operaties. Je laat een graafmachine ook geen archeologisch onderzoek doen. De computer is maar een deel van het geheel, pas als je alle processen op orde hebt, kan de computer iets toevoegen.” 