



Guus Schreiber, hoogleraar Intelligente Informatiesystemen aan de Vrije Universiteit, doceert een groep promovendi. Samen met collega Frank van Harmelen leidt Schreiber de informele en succesvolle Semantic Web Group. Schreiber werkt onder meer samen met het Tropenmuseum. 'Als iets ons onderzoek typeert, is het dat we nieuwe theoretische wegen onderzoeken en toch heel pragmatisch blijven'

Reportage VU Amsterdam pioniert in zelfdenkend internet

Spin in het Semantische Web

De verwachtingen zijn hooggespannen: het Semantische Web is het internet van de nabije toekomst. Nog even en dan kunnen computers intelligent gegevens verzamelen en ordenen, om er zelfstandig onderzoek mee te doen en ontdekkingen te verrichten. Aan de VU in Amsterdam werkt een club wetenschappers aan dat visioen. Maar komt dat 'Web 3.0' er ooit? En hoe zit het met de privacy?

Tekst Peter Pijls
Foto's Irma Bulkens

Peter Mika is de *hot young rising star* van het Nederlandse onderzoek naar het nieuwe internet. De Hongaar schittert met internationale prijzen, is door de internetgigant Yahoo! weggekocht nog voordat hij zijn promotie heeft afgerond en wordt door zijn professoren aan de Vrije Universiteit (VU) in Amsterdam op handen gedragen. Mika is, aldus de internationale vakpers, 'een van de meest opmerkelijke Semantische Web-onderzoekers ter wereld'.

Zelf noemt hij zich 'gewoon een praktische jongen'. Hij legt de achtergronden van dat Semantische Web uit. 'In mijn promotie-onderzoek wil ik aantonen dat je via internet straks alle gegevens kunt analyseren en wellicht verbinden.' Dat is ook de kern van het Semantische Web: computers moeten binnen afzienbare tijd zelf onderzoek kunnen doen door gegevens die via internet beschik-

baar zijn, aan elkaar te koppelen. Zo kunnen bijvoorbeeld nieuwe medicijnen worden ontdekt, kunnen consumenten hun inkopen laten doen of opdrachten laten uitvoeren en kunnen museumcollecties volautomatisch worden ontsloten.

Mika: 'Dit is ons visioen: computers zelf ontdekkingen laten doen.' Professor Frank van Harmelen, hoogleraar Kunstmatige Intelligentie aan de VU, vult aan: 'Veel informatie is nu al vrij beschikbaar op het web, alleen niet in een format waar computers iets mee kunnen. Het Semantische Web maakt het mogelijk dat computers die informatie gebruiken om ons te helpen te zoeken en te vinden.'

Van Harmelen leidt samen met zijn collega professor Guus Schreiber, hoogleraar Intelligente Informatiesystemen, de informele *Semantic Web Group* aan de Vrije Universiteit van Amsterdam. Met financiële steun van de Europese Unie en NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) brainstormt

die groep over de ontwikkeling van deze nieuwste en revolutionaire toepassing van het internet.

De VU behoort tot de vijf beste instituten ter wereld op dit gebied. In toekomstvisioenen is het Semantische Web – ook wel Web 3.0 genoemd – de droom van veel wetenschappers: computers die zelf ontdekkingen doen door enorme hoeveelheden data zelfstandig te analyseren en te combineren. Internet als gigantisch brein en alwetende machine. Het was de uitvinder van internet zelf die het idee zo naar buiten heeft gebracht. Tim Berners-Lee, de internetgoeroe die in 1991 het World Wide Web lanceerde, is nu de drijvende kracht achter het Semantische Web. Vijf jaar geleden voorspelde hij in een veelbesproken artikel in *Scientific American* dat websurfers met semantische zoektechnologieën hun dagelijks leven zullen kunnen organiseren. En dat al in een nabije toekomst. Zover is het nog niet, nuanceert Van Harmelen. Maar hij ontkent niet dat de mogelijkheden in het verschiet liggen.



De tweewekelijkse bijeenkomst van de Semantic Web Group. Er zijn zeventien promovendi en postdocs aanwezig. Oost-Europa is met Bulgarije, Hongarije, Oekraïne en Macedonië goed vertegenwoordigd →

Gang van de Faculteit der Exakte Wetenschappen aan de VU. De VU behoort tot de vijf beste instituten ter wereld op het gebied van het Semantisch Web

Wat semantische zoekmachines doen, gaat veel verder dan Google en Yahoo!. Die doorzoeken het internet en geven hun resultaten weer in plaatjes en tekst in de HTML-computertaal. De semantische aanpak verandert deze in een web van aan elkaar gekoppelde gegevens. Web 3.0 kan beschikbare informatie daardoor niet alleen beter selecteren en combineren, maar zelfs persoonlijke wensen begrijpen. Daarvoor is een nieuwe computertaal ontwikkeld, OWL, de 'HTML' van het Semantische Web.

De Vrije Universiteit speelde bij het ontwerpen van die taal een belangrijke rol. 'Dat was voor de VU instrumenteel bij het verkrijgen van onze leidende positie in het wereldwijde onderzoek naar het Semantische Web,' zegt Van Harmelen, een jongensachtige, beweeglijke veertiger. 'Het was een cruciaal stuk technologie.'

De nieuwe computertaal die het Semantische Web ondersteunt, is meer dan een persoonlijk succes van de Amsterdamse hoogleraren. Ze is nodig, omdat het huidige

internet volgens de onderzoeker aan het plafond van zijn kunnen zit. 'De druk van de enorme zee aan webpagina's wordt zo groot, dat die zijn waarde verliezen als we geen betere ondersteunende hulpmiddelen verzinnen. Het Semantische Web moet er komen. Het gaat om noodzakelijke ontsluiting. Zoekmachines als Google worden niet meer beter.'

Van Harmelen bemerkte aanvankelijk een afwachtende houding bij de grote zoekmachines als Yahoo! en Google. Mails over toepassingen van het Semantische Web werden niet beantwoord. Maar hij ziet beweging: 'Ze zijn ermee bezig. Een van onze promovendi gaat naar Yahoo!. Ilse in Nederland experimenteert ermee. Een paar afstudeerders van ons hebben bij Ilse een project gedaan om semantische technieken toe te passen. En die mensen zijn meteen weggekocht.'

Picasso

Op een maandagmiddag eind oktober houdt de Semantic Website Group haar

tweewekelijkse bijeenkomst. Professor Guus Schreiber illustreert met een beamer hoe een semantisch programma kunstcollecties intelligenter doorzoekt met zoek- en informatievragen. Een vraag naar 'Picasso' levert meer op dan de gebruikelijke Googlebrij. Het semantische programma onderscheidt afzonderlijke categorieën: objecten gemaakt door Picasso uit verschillende periodes of materialen, schilderijen waarop de kunstenaar zelf is afgebeeld of gereleerde werken gemaakt door andere kubistische kunstenaars. Enkele weken later wint het project de eerste prijs op de 5th International Semantic Web Conference in Athens, Amerika.

Bij de presentatie zijn zeventien promovendi en postdocs aanwezig uit acht verschillende landen. Oost-Europa is met Bulgarije, Hongarije, Oekraïne en Macedonië goed vertegenwoordigd. De jonge Hongaarse promovendus Mika zit helemaal achter in het lokaal en plaatst zo nu en dan kanttekeningen. Af en toe schiet hij in de lach.





← Fran
Kun
enon
dat d
geen
verz

Con
Wa
'Er is
B

Profe
vend
groel
de be
kwan
in di
dier
een
een
de ac
een l
plan
een l
ken?

De
de N
perta
ken.
Van
in 20
The
man
hij d
Web

← Frank van Harmelen, hoogleraar Kunstmatige Intelligentie. 'De druk van de enorme zee aan webpagina's wordt zo groot, dat die zijn waarde gaat verliezen als we geen betere ondersteunende hulpmiddelen verzinnen'



Computerlokaal van de Faculteit der Exacte Wetenschappen van de VU. Van Harmelen: 'Er is nog altijd geen supercomputer die in de buurt komt van wat wij tussen onze oren hebben'

Professor Van Harmelen noemt de promovendi 'het levensbloed van de onderzoeksgroep'. Hij merkt op dat de laatste tien jaar de beste aio's en postdocs uit het buitenland kwamen: 'De Nederlandse promovendus is in dit vakgebied een tijd lang een zeldzaam dier geweest. De internethype heeft daar een rol in gespeeld. Als je de voorkant van een toetsenbord kon onderscheiden van de achterkant, dan kon je in de IT-wereld al een baan met een lease-auto en een bonusplan krijgen. Dus waarom zou je dan voor een laag loon aan een universiteit gaan werken?'

De VU wist het gebrek aan gemotiveerde Nederlanders goed te maken door supertalenten uit Oost-Europa aan te trekken. Peter Mika is 'een bijzonder geval', zegt Van Harmelen. De 28-jarige Hongaar won in 2004 in Hiroshima de eerste prijs tijdens *The Semantic Web Challenge* met een semantische applicatie. Vorig jaar scoorde hij de *Best Paper Award* op een Semantic Web Conferentie in Ierland. De vakpers

noemt hem een van de meest opmerkelijke Semantische Web-onderzoekers ter wereld.

Onder de motorkap

De wetenschappelijke successen en de internationale roem die de VU en haar medewerkers en promovendi oogsten, staan haaks op de grote onbekendheid van het Semantische Web en zijn toepassingen. Van Harmelen: 'Dat is de normale cyclus van een wild idee dat begint in onderzoekslabs, waar onderzoekers zich er met subsidie een paar jaar drukover maken. Vervolgens worden kleine innovatieve bedrijven wakker. Dat stadium is nu bereikt, vooral in Amerika. Ze gaan producten ontwikkelen, ze beginnen winst te maken en trekken de aandacht van de grote bedrijven. Die kopen de kleintjes op en dan wordt het product zichtbaar voor het grote publiek. Het semantische web is langs die lijn op reis. Een andere reden waarom het grote publiek het niet kent, is dat de technologie onder de motorkap zit. Je ziet het niet direct.'

De Semantic Web Group voelt niettemin wel druk om met concrete en vooral publieke resultaten te komen. De grote subsidiegevers, de Europese Unie en NWO, dringen daarop aan. Frank van Harmelen noemt de prestatiedruk 'vervelend'. Hij signaleert een hijgerige cultuur: 'De redenering is: je doet nu al zo lang onderzoek en we zien maar geen praktische toepassingen. Gaat het eigenlijk wel goed met dat onderzoek van jou? Die stemming heerst bij de industrie, maar ook wel bij de fondsenverstrekkers in Europa en bij NWO. Ze geven je het gevoel dat de horizon betrekkelijk kort is. Dat je snel met huiskamerproducten op de proppen moet komen.'

Ook zijn collega professor Schreiber wordt soms onrustig van de stille hype rond het Semantische Web: 'Je moet gisteren resultaten behaald hebben. Men heeft onrealistische verwachtingen. Soms gaan onderzoekers zich daardoor beter voordoen dan ze zijn. Daar heb ik een ontzettende hekel aan. Als iets ons onderzoek typeert, is het dat we



Presentatie tijdens een bijeenkomst van de
Semantic Web Group. Van Harmelen: 'De
Nederlandse promovendus is in dit
vakgebied een tijd lang een zeldzaam dier
geweest' →

De serverruimte van de Faculteit der Exacte
Wetenschappen. Van Harmelen: 'De
kunstmatige intelligentie, met al z'n
creativiteit en pioniersgeest, heeft ook een
heel vervelende geschiedenis van mensen
met een te grote mond'

nieuwe theoretische wegen onderzoeken en toch heel pragmatisch blijven.'

Praktische toepassingen

Intussen werken de onderzoekers op de VU wél aan praktische toepassingen. Van Harmelen doet een project met British Telecom om het intranet van het bedrijf semantisch te ontsluiten. Hij wijst erop dat de navigatie van e-commerce sites als Amazon al deels semantisch werkt. 'Voorspellen is moeilijk, helemaal op het web. Maar binnen een paar jaar zouden zulke toepassingen het grote publiek moeten bereiken.'

Schreiber is voorzichtig: 'Ik probeer aan *expectation management* te doen. Wat kunnen we reëel verwachten? Als we succesvol zijn, zal er binnen een paar jaar op het web een aantal semantische eilanden ontstaan, waar mensen thuis daadwerkelijk iets mee kunnen achter hun pc. Dat zou me zeer tevreden stemmen. Denk aan publieke informatie, zoals bibliotheken, erfgoed, geneeskunde en biotechnologie, waar heel veel

informatie is die gedeeld moet worden.'

Schreiber wijst subsidieverstrekkers op heel praktische toepassingen van zijn onderzoek. Hij werkt samen met het Rijksmuseum voor Volkenkunde en het Tropenmuseum om hun collecties met semantische technieken te ontsluiten voor webgebruikers. Het heeft lang geduurd, zegt Schreiber, om mensen uit de culturele wereld warm te krijgen voor het onderzoek: 'Bèta meets alfa, dat idee. In die wereld vindt men het een beetje eng wat we doen. Er is een sociale barrière, maar ze zien het wel als hun missie om de collecties te ontsluiten. Je moet een proces genereren waarin steeds meer mensen gaan meewerken. Dat lukt door een leuke toegevoegde waarde te laten zien. Dan kan het snel gaan. Maar bij de EU en de Nederlandse overheid wordt altijd heel erg de nadruk gelegd op het economische belang.'

Internetpionier

Tim Berners-Lee, de internetpionier en drijvende kracht achter het Semantische Web,

heeft zelf in niet geringe mate bijgedragen aan de hooggespannen verwachtingen rond het Semantische Web, zeggen de Amsterdamse onderzoekers.

Met zijn voorspelling dat computergebruikers binnenkort hun leven semantisch kunnen organiseren, ging hij te ver, vinden Schreiber en Van Harmelen. Ze weerleggen Berners-Lee's opvattingen dat het Semantische Web een wereldomvattend brein kan worden. Van Harmelen: 'Met dat soort uitspraken balanceert hij op de rand. Ik ben daar niet gelukkig mee. De kunstmatige intelligentie, met al z'n creativiteit en pioniersgeest, heeft ook een heel vervelende geschiedenis van mensen met een te grote mond die hun pretenties niet hebben waargemaakt. In de jaren zestig riep Herb Simon al: over vijf jaar zijn we wereldkampioen schaken met computers. De kunstmatige intelligentie heeft een heel verleden van opgeblazen verwachtingen. Neem converseren met computers in natuurlijke taal. Spraakherkenningsoftware, oké, vooruit, maar dat





Terrein van de Faculteit der Exacte Wetenschappen van de Vrije Universiteit in Amsterdam

is nog heel iets anders dan een intelligent gesprek voeren met je computer. *Forget it.* Van die opgeblazen verwachtingen hebben we enorm veel last.'

Van Harmelen wijst daarom ook graag op de beperkingen van internet en het Semantische Web: 'Er is nog altijd geen supercomputer die in de buurt komt van wat wij tussen onze oren hebben. Puur alleen al aan rauwe hardware komen we digitaal zwaar tekort. Maar we hebben ook nog geen begin van een goed begrip over hoe cognitieve processen werken.

Hoe komt het dat ik jou morgen in het donker feilloos herken nu ik je onder deze belichting voor het eerst zie? Er is geen computer die bij dat soort processen in de buurt komt. In populaire lezingen gebruik ik de metafoer van het Semantische Web dat een omgevallen boekenplank, het World Wide Web, verandert in een georganiseerde kennisbank waarin je de dingen een beetje beter aan elkaar kan knopen. Maar dat is geen intelligent wereldomspannend brein en dat is daar ook niet voor nodig. Van die misvatting wil ik af.'

Privacy

Critici van het Semantische Web vrezen voor de privacy van de computergebruiker. Daar kan Van Harmelen zich wel in vinden. 'De privacykwestie is een terechte zorg. Volgens mij heeft nog niemand het goede antwoord. Misschien is het opgeven van privacy een prijs die je moet willen betalen. Je kunt niet verwachten dat websites voor jou betere zoekresultaten leveren als je niet wilt vertellen waar je naar op zoek bent. En vertellen waar je naar op zoek bent, betekent iets van je privacy weggeven. Daar zullen we in een nieuwe informatiewereld evenwicht in moeten vinden. Maar nu is dat ook al het geval. Met de bonuskaart van Albert Heijn geef je ook je privacy op.'

De Hongaarse promovendus Peter Mika, binnenkort naar Yahoo!, is daar minder ongerust over.

'Er is een belang om je identiteit te beschermen. Maar er is ook een algemeen belang om zoekresultaten met elkaar te verbinden. Dat is een delicaat evenwicht. We zullen privacy inderdaad opnieuw moeten definiëren. Zoekmachines weten nu al meer over ons dan we willen toegeven. Ze weten wat jij zoekt en ze slaan het op. Zo kunnen ze gericht adverteren. Daar leven ze van. Gmail, het mailprogramma van Google, plaatst nu al advertenties op jouw mailbox gebaseerd op de inhoud van je mails, hoe- wel heel veel mensen dat ongetwijfeld niet

weten. Maar privacy is iets persoonlijks. Zelf hoor ik bij een generatie die bereid is wat privacy op te geven. Ik weet wat er mogelijk is op het web en ik ben er niet paranoia over. Wat dat betreft zit ik in de *'giving up mode'*. Het National Security Agency in de VS monitort al het inkomende en uitgaande e-mailverkeer. Wat moet je daartegen doen? Er is wel oppositie. Je hebt altijd mensen die petities indienen. Maar ik vind de innovatie van het Semantische Web belangrijker. *I want to move ahead.*'

Het eerste faxapparaat

In zijn promotieonderzoek stelt Mika dat sociale netwerken in een semantische toekomst alle gegevens op het internet met elkaar kunnen verbinden. Een ambitieuze gedachte, beaamt hij. Droog: 'Mijn theorieën stuiten op nogal wat weerstand.' Niettemin gelooft hij heilig in het Semantische Web. Hij vergelijkt het in zijn proefschrift met de introductie van de fax. Het eerste faxapparaat was duur en nauwelijks bruikbaar omdat niemand anders er een had. Hoe meer faxapparaten er verkocht werden, hoe bruikbaarder ze werden.

In een verlaten collegezaal van de Vrije Universiteit maakt de Hongaar zijn punt. Het is een opwindend moment om te promoveren op het Semantische Web, stelt hij: 'Toen ik begon als promovendus was het onderwerp eerder riskant dan aantrekkelijk. Nu hebben we een kritisch punt bereikt. Er is genoeg research gedaan, er zijn nieuwe computertalen ontwikkeld. Het gros van het voorbereidende werk is gedaan. *This should be the time.* Maar we moeten het eens worden over nieuwe computertalen, gedeelde concepten en symbolen die iedereen begrijpt. Dat maakt het Semantische Web nog moeilijk te verkopen.'

Zal het web uiteindelijk gaan functioneren als een gigantisch brein?

Peter Mika beantwoordt de vraag niet direct. 'In mijn proefschrift draai ik tot op zekere hoogte om de hete brij heen. Want het Semantische Web is problematisch als het gaat om het verbinden persoonlijke gegevens. Niemand heeft daar op dit moment een antwoord op.'



Zie ook: www.cs.vu/~frankh/
 en www.cs.vu/~guus/