



Het web moet slim en sociaal worden

Van onze verslaggever Bas Timmers

gepubliceerd op 14 december 2007 12:47 , bijgewerkt op 14 december 2007 13:26

AMSTERDAM - Meer dan 1.200 experts kwamen 11 en 12 december bijeen op de Le Web 2007-conferentie in Parijs om de toekomst van het web te bespreken. Het woord 'sociaal' stond hoog op de agenda. Want het web moet slimmer en persoonlijker worden.



Google speurt naar een verzameling letters die samen een woord vormen.

Wat er eigenlijk mis is met Google? Frank van Harmelen zucht even, omdat hij weet dat gebruikers zo tevreden zijn over de zoekmachine. 'Bij de zoekresultaten zit veel bagger.', vertelt de hoogleraar kunstmatige intelligentie aan de VU in Amsterdam. 'Zoek je bijvoorbeeld op Harmelen, krijg je zowel artikelen over mij als over het dorp met die naam.'

Poesjes

Want een zoekmachine doet niets meer dan op het web speuren naar een verzameling letters die samen een woord vormen. Het heeft geen benul wat die tekst betekent, laat staan wat de gebruiker precies hoopt te vinden. Van Harmelen's 8-jarige dochter moest onlangs voor school een werkstuk maken en gaf de zoekopdracht 'plaatjes van poesjes'. Dat werd een ongewilde educatieve ervaring voor het meisje.

Momenteel lossen eenvoudige 'parental control'-filters dat probleem overigens redelijk op, al filteren die weer te veel resultaten uit. 'Je wilt dus een betere selectie door Google, men moet begrijpen dat een basisscholier iets anders zoekt dan een volwassene. Maar Google zit aan zijn top, het wordt nauwelijks meer beter. Dus dan moet je het over een heel andere boeg gaan gooien.'

Semantisch web

Van Harmelen is niet de enige die vindt dat de koers zou moeten gaan richting een 'semantisch web'. Tim Berners-Lee, de man die het world wide web eind jaren tachtig uitvond, schreef in 2001 een artikel over het onderwerp in Scientific American. Het stuk vormde het startsein voor het door Berners-Lee geleide W3C, dat de ontwikkeling van het internet bewaakt, om standaarden (zie onder dit artikel) te ontwerpen voor het semantisch web.

Dat semantisch web is in feite een groep gestructureerde woordenboeken, die een computer vertellen wat de betekenis van bepaalde begrippen is. Daardoor moet programmatuur veel slimmer worden. Hij moet weten dat Harmelen zowel een dorp als een persoon is en welke van de twee betekenissen in deze context van toepassing is. Als iemand zoekt naar 'dessert', is het handig als de computer ook speurt naar 'toetje'. En wie het web afgraast naar een nieuw type videocamera, zou automatisch tips moeten krijgen over apparaten die dezelfde mogelijkheden hebben. 'Een goede verkoper doet dat immers ook', weet Van Harmelen. 'Amazon doet het al een beetje. Als je een boek koopt, krijg je meteen tips om andere titels aan te schaffen. Om die reden zie ik ook de grootste mogelijkheden in de e-commerce.'

Heilige graal

Die commerciële kansen zijn voor sommigen aanleiding om een semantisch web (soms overigens web 3.0 genoemd) te bestempelen tot de heilige graal van het internet. De hoogleraar waakt echter voor overspannen verwachtingen. 'IT heeft een lange historie van beloftes die niet waargemaakt konden worden.'

Toch is de laatste jaren een stevig fundament komen te liggen onder het semantische web. Miljoenen gebruikers geven steekwoorden (tags) mee aan bijvoorbeeld de foto's, video's en ervaringen die ze plaatsen op sites als YouTube en Flickr. Ook zijn er programma's die volledig automatisch steekwoorden genereren uit teksten en databases. Dat levert een schat aan extra informatie (in vaktaal metadata) op.

Door gestructureerde woordenboeken te gebruiken, kan het web straks die steekwoorden aan elkaar spinnen. Zoekmachines moeten weten dat 'Mokum' en 'Amsterdam' hetzelfde betekenen, en dat Amsterdam in Nederland ligt. Voor een mens voor de hand liggend, voor een computer niet. Aan de woordenboeken én de programmatuur om alle informatie aan elkaar te knopen wordt nu gewerkt. 'Ik was pas nog op bezoek bij Microsoft in Redmond om hen te helpen met het maken van een onderzoeksprogramma op dit gebied', vertelt Van Harmelen. 'Zij houden de ontwikkelingen op dit terrein scherp in de gaten.'

Weerbarstige materie

De hoogleraar is dagelijks bezig om de motor achter het web te verbeteren. Maar het blijft een weerbarstige materie om computers verbanden te laten leggen, iets wat mensen automatisch doen. Al boeken Van Harmelen en collega's successen op kleine, goed afgebakende terreinen.

De Vrije Universiteit, volgens Van Harmelen één van de vijf toonaangevende instituten op dit terrein, legde bijvoorbeeld een semantisch woordenboek aan over de collectie van het Rijksmuseum. Daardoor snapt de browser nu het verschil tussen schilderijen door Van Gogh, of schilderijen met Van Gogh er op. En het lukt de browser om schilderijen in verband te brengen met informatie uit een andere website over de periode waarin Van Gogh leefde, en welke andere schilders toen actief waren. 'Kijk', zegt Van Harmelen, 'da's al een hoop informatiever dan alleen een plat Google-lijstje met hits, waar je het zelf verder mag uitzoeken. En dit kan alleen maar doordat de computer nu iets 'weet' over schilderijen, en over Van Gogh, doordat er van die gestructureerde woordenboeken achter zitten.'

Maar het komt nog niet in de buurt van het visioen van Berners-Lee. Die droomde in 2001 al dat je in 'gewone-mensen-taal' aan de computer een vraag kon stellen als: zoek een restaurant in een straal van tien kilometer, waar ik Mexicaans kan eten voor een bescheiden budget en geef meteen de routebeschrijving. Oh ja, en kijk even of mijn vrienden misschien een restaurant aanraden, zo was (vrij vertaald) de wens die Berners-Lee daar twee weken geleden nog aan toevoegde op zijn weblog.

Kunstmatige intelligentie

Christine Karman, specialist in kunstmatige intelligentie en winnaar van de ICT publieksprijs 2005, gelooft echter heilig dat het ooit realiteit wordt. 'Dat pc's op een dag intelligenter worden dan mensen, staat vast. Wie tien jaar geleden had geroepen dat een computer beter zou kunnen schaken dan een mens, werd ook uitgelachen.'

Maar of de grote doorbraak komt via het semantische web, betwijfelt Karman. 'Ik geloof niet dat je alle relaties tussen woorden kunt vastleggen. Zo'n web is niet te onderhouden en wordt zo complex, dat het eerder minder overzichtelijk wordt.'

Ze ziet meer in computers die zelf ontdekken welke begrippen met elkaar te maken hebben. Of, zoals David Weinberger het beschreef in zijn boek Everything Is Miscellaneous: 'De computer ziet dat veel foto's op Flickr van de Golden Gate brug zowel een tag San Francisco als Californië hebben. En dat het steekwoord Californië ook vaak samen gebruikt wordt met Los Angeles en San Diego. Dan kan de pc tot de voorlopige conclusie komen dat Californië de wortel is van een boom, waar Los Angeles en San Francisco de takken van zijn.'

Zelf-lerende agent

Karman noemt zulke programma's 'agents'. In feite was ze haar tijd ver vooruit toen ze met het door haar opgerichte bedrijf Tryllian in 2001 zo'n intelligent zelf-lerende agent, de Gossip, introduceerde. Dat stukje software kon met andere Gossips communiceren om te leren en zo met suggesties terug te komen richting gebruiker. Het idee sloeg niet aan, maar Karman pikt het binnenkort weer op. 'Want iedereen zit nu te chatten en te skype, dat is een ideale omgeving om

dit soort applicaties te schrijven.'

Daarmee hint ze op misschien wel de grootste belofte van het semantische web.

Internetgebruikers laten tegenwoordig namelijk een schat van informatie achter op het web. Ze boeken hun vakanties, kopen boeken en cd's, schrijven recensies van films, voegen steekwoorden toe aan foto's en bewaren hun favoriete pagina's. Als computers die data aan elkaar gaan knopen, komen ze met veel bruikbaarere tips. Al raak je dan wel het gevoelige punt van privacy. 'Maar die discussie is schuivend', vertelt Van Harmelen. 'Een pc kan nu eenmaal betere kunstjes naar mate hij meer over je weet. Daar ligt een groot potentieel. We wennen er steeds meer aan dat we een stukje privacy inleveren voor betere diensten. Kun je je de discussie over de Bonuskaart van Albert Heijn nog herinneren? Waar precies de balans zal komen te liggen weet niemand.'

Schat aan informatie

Zelf heeft hij weinig moeite met het prijs geven van sommige privacygevoelige informatie. Op het web heeft Van Harmelen onder andere zijn agenda, huisadres, foto's en favoriete boeken staan. Als iedereen dat zou doen (en die trend is gaande) levert het een schat aan informatie op. En wordt Google, of de opvolger van dat nu almachtige bedrijf, eindelijk een stuk socialer. Of zoals Karman het zegt: 'Een intelligente browser die zelf leert en zelf op zoek gaat.'

Frank van Harmelen (1960) is hoogleraar kunstmatige intelligentie aan de Vrije Universiteit in Amsterdam. Hij schreef onder andere het boek 'The Semantic Web Primer'. Christine Karman is al sinds de jaren tachtig bezig met praktische toepassingen op het gebied van kunstmatige intelligentie. Ze werkt sinds 1995 als technologie-expert, richtte in 1998 Tryllian op en was tot voor kort directrice van Izecom (beveiliging van e-mail).

De technologie achter het semantisch web

De steekwoorden (tags) bij documenten worden in de zogenaamde XML-taal beschreven. Een webpagina kan bijvoorbeeld de tags 'auto, verkoop, Volkswagen' bevatten. Een computer heeft echter geen idee wat een auto is. Daarvoor is RDF, één laag boven XML, uitgevonden: dat beschrijft bijvoorbeeld 'auto kan rijden' en 'auto heeft vier wielen'. Dat voegt betekenis toe aan woorden. Maar iemand kan in plaats van 'auto' natuurlijk ook het woord 'voertuig' of 'automobiel' gebruiken. OWL, de bovenste laag van het semantische web, zorgt dat computers die verschillende benamingen voor hetzelfde onderwerp aan elkaar kunnen koppelen. En ook kan iemand OWL gebruiken om te vertellen dat een auto onderdeel is van de groep 'transportmiddelen', waar trein en vliegtuig ook toe behoren.

Verder heeft het semantisch web (onder andere) ook een sociale component. De Friend Of A Friend-technologie (FOAF) vertelt wie met wie bevriend is. Dat kan ertoe leiden dat bij de zoekvraag 'ik wil een spannend boek lezen' tips van het vriendennetwerk veel belangrijker zijn dan tips van mensen die niet met de zoeker verbonden zijn.

Links:

[Demo van de slimme kunst-browser](#)

[Artikel Berners-Lee in Scientific American](#)

[Blog Berners-Lee](#)

[Le Web](#)

[Boek David Weinberger](#)

[Open Directory](#) (index van websites, bijgehouden door vrijwilligers)

[Collexis](#) (Nederlands bedrijf, actief met semantisch web-technieken)

[Aduna](#) (Nederlands bedrijf, actief met semantisch web-technieken)

 [PRINTVERSIE](#) | [VERSTUREN](#) | [RSS](#) | [VOLGEND ARTIKEL](#)