

Je zoekt iets op het web en krijgt een berg ongeordende data. Dat kan beter. Met het nieuwe, 'betekenisvolle' zoeken krijg je informatie die jij echt zoekt. En meer. Guus Schreiber legt uit.

Onder je slimme vriendjes hoef je het waarschijnlijk niet te proberen, maar bij je modale kennissen kun je er behoorlijk indruk mee maken: even de term 'semantisch webbing' laten vallen. Of over 'Web 3.0' beginnen. Geheid dat er respectvolle en licht verwarde blikken je kant opkomen. 'Hüh, wat zeg je nou?'

Enfin, daarna wordt het moeilijker, want dan moet je het dus uitleggen. Dat er een nieuwe tijd is aangebroken qua gebruik van het net, dat het Oude Zoeken op internet vervangen gaat worden door het Nieuwe Zoeken, omdat de brij van digitale informatie die over het World Wide Web zweeft, in de toekomst geheel anders georganiseerd gaat worden.

Dat klinkt weer heel simpel, wat jammer is voor de verpletterende indruk die je net hebt gemaakt, maar VU-hoogleraar Guus Schreiber, die er alles van weet, wil dat ook graag, dat er niet te moeilijk over wordt gedaan.

Want waar hebben we het over? Dat we van dom zoeken, met een onderwerp ingetikt in het zoekveld, voortaan 'betekenisvol', zeg maar 'semantisch' gaan zoeken. Bij het oude domme zoeken krijg je bij een ingetikt woord een rij documenten die allemaal iets zeggen over jouw term of onderwerp en die, ruwweg, zijn geordend op populariteit: wie het hoogst staat, wordt het meest gezocht en dat krijg je dus als (be)zoeker het eerst aangeboden. Maar veel succes verder, bij het filteren van die reeks aan elkaar gekoppelde documenten of *linked documents*.

Picasso

Bij het betekenisvolle zoeken, krijg je het antwoord op je vraag niet als bulkgoed aangeboden, maar in verschillende, samenhangende categorieën. Een cliché-voorbeeld is het zoeken naar informatie over de schilder Picasso. Normaal krijg je een hoop ongeordende informatie met 'Picasso' erin, maar bij het nieuwe zoeken verschijnt: Picasso de man, zijn werk, zijn netwerk, zijn vrienden, vrouwen, noem maar op. De informatie is én geordend én tegelijk zinvol geïntegreerd. En het leuke is: je krijgt soms ook geheel nieuwe kennis aangeboden. Dat de Picasso ook een auto-type is, weten we wel, maar wie wist, zoals een aparte categorie meldt, dat het ook een steensoort is?

Guus Schreiber wist het niet, verklapt hij op zijn werkkamer in het Amsterdamse VU-gebouw. Hij is hoogleraar Web & Media en leidt samen met collega Frank van Harmelen de informele Se-



Guus Schreiber: 'We gaan van een zoekstelsel dat documenten ophoest, naar een stelsel van gekoppelde data.'

Web 3.0: meer vinden dan je vraagt

mantic Web Group. Met hun vijftig medewerkers – onder wie voor Informatica begrippen relatief veel vrouwen – vormen zij een van de leidende onderzoeksgroepen die wereldwijd op dit onderzoeksgebied actief zijn. Hij werkt samen met Sir Tim Berners-Lee, de internetgoeroe die in 1989 het World Wide Web bedacht, en van de VU vorig jaar een eredoctoraat ontving.

Maar Schreiber doet dus liever niet al te ingewikkeld over de internetrevolutie die gaande is. Hij zegt eenvoudig: 'We gaan van een zoekstelsel dat documenten ophoest, naar een stelsel van gekoppelde data.' *Linked data* dus.

Vervoersstromen

Wat is het verschil? Schreiber: 'Neem nou de gegevens over de vervoersstromen in Nederland die je op het net kunt vinden. Dat kun je moeilijk 'documenten' noemen waar je eens lekker voor

gaat zitten. Maar die informatie is wel heel interessant, vooral als je de gegevens, geheel of gedeeltelijk, combineert met andere informatie over files of vrachtwagens, op een manier die niet alleen jouw vraag beantwoordt, maar sterker nog: tot nieuwe inzichten leidt en jou nieuwe informatie verschaft.'

Iets vragen, en dan een antwoord krijgen dat meer geeft dan je vroeg. 'Dat is het mooie van deze nieuwe mogelijkheden, dat je verrijkte kennis aangeboden krijgt', zegt Schreiber. Nog deftiger wordt het aangeduid met de Engelse term *serendipity*, wat betekent: het vinden van iets onverwachts en bruikbaar terwijl je op zoek bent naar iets totaal anders. Dat wordt dus een behoorlijk intelligente manier van zoeken.

De relativerende professor wil daarbij meteen een snel opstekend misverstand in de kiem smoren. 'Dat maakt van de computer géén denkende machi-

ne of een megabrein, zoals mensen direct roepen.' Volgens Schreiber kun je een menselijk brein niet vergelijken met een machine, ook niet met een slimme computer. 'Je laat een mens toch ook niet hardlopen tegen een auto? Het zijn volstrekt verschillende grootheden.'

Geen romantische techno-vergezichten dus. Wat niet wil zeggen dat er geen prachtige dingen staan te gebeuren met dat nieuwe web-weten. Maar om anders te kunnen zoeken, moet de informatie ook anders toegankelijk worden gemaakt, en dat wordt nog een behoorlijk klus. In het 'oude' systeem zet de browser op je computer de HTML-'syntax' (het coderingssysteem of de 'taal') om in een pagina. Wat er op die pagina staat, wat het allemaal betekent, weet de computer niet.

Bij het nieuwe zoeken moet om te beginnen alle informatie zodanig worden

gelabeld, dat universeel wordt begrepen waar dat label voor staat. RDF is de taal die het Web hiervoor gebruikt. Als gebruiker zie je daar niets van, maar als je bijvoorbeeld op Facebook op de 'Like' button klikt, wordt er een stukje RDF aangemaakt. Vervolgens kan die informatie beter worden begrepen en toegepast worden op de vraag die jij, met je persoon-specifieke vragen en behoeften, stelt.

Musea

Zitten we dan over een paar jaar massaal op het semantische web? Nee. Alleen al het coderen en koppelen van wereldwijde digitale informatie is zo'n bewerkelijke klus, dat het nog wel even zal duren voordat het nieuwe zoeken de standaard is geworden. 'Er zullen steeds meer eilandjes van samenhangende, vaak gevisualiseerde, data komen en dat eilandrijk zal zich steeds verder uitbreiden', schetst Schreiber de toekomst. Een stille, geleidelijke ontwikkeling dus, geen big bang.

Schreiber: 'Het voordeel van deze *linked data*-ontwikkeling is dat het heel praktijkgericht is. We zijn bijvoorbeeld al een tijdje bezig collecties van diverse Nederlandse musea te inventariseren

Het leuke is: je krijgt geheel nieuwe kennis aangeboden

en te ordenen, zodat de schat aan culturele informatie die daar is, voor iedereen beter toegankelijk wordt en nieuwe en interessantere inzichten biedt.'

Wie in het Amsterdamse Rijksmuseum met behulp van de VU-demonstrator al eens een 'gepersonaliseerde rondleiding' heeft gedaan – ingericht dus op persoonlijke voorkeur – in plaats van gedachte-loos te hebben rondgedwaald door de vele zalen, heeft z'n eerste vrucht van 'semantisch zoeken' al binnen. 'Het Rijksmuseum heeft als handelsmerk kwaliteit, maar is tegelijk voor ons onderzoekers een heel vooruitstrevende partner.'

Publieksvriendelijk zijn, dat staat bij Schreiber en zijn collega's voorop. Een ander belangrijk doel is daarbij de burger gedienschtig te zijn. 'President Obama heeft in die zin al een statement gemaakt door te verklaren dat hij aanhanger is van 'open government data', waarbij alle, niet persoonlijke, overheidsgegevens openbaar beschikbaar worden gesteld.' Burgers kunnen dan zelf informatie combineren en zijn daarvoor niet langer afhankelijk van ambtenaren. Schreiber: 'Ook ik wil de openheid van het net met hand en tand verdedigen ten behoeve van onze kennisdemocratie.'

Om het net voor iedereen nog mooier, beter en persoonlijker te maken hoopt Schreiber dat er snel een onderzoeksgedigitaliseerde moderne leven voorwerp van studie wordt. 'We hebben daarbij hard vrouwen nodig, we zoeken multitaskers, mensen met verstand van lifestyle en psychologie én die wat weten van ict – sociale technen dus.'