

Het slimme broertje van Google

Volgens VU-wetenschappers zitten zoekmachines als Google op hun top: de kwaliteit van de zoekresultaten zal niet opzienbarend verbeteren. Tenzij het semantisch web een succes wordt. Dan weet de zoekmachine net als wij dat Den Bosch en 's-Hertogenbosch dezelfde plaatsen zijn. Twee hoogleraren kijken terug én vooruit.

door **Anouck Vrouwe** 2001 ***The Semantic Web***

Tim Berners Lee, de man die in 1991 het World Wide Web begon, schrijft een artikel in Scientific American over de toekomst van het web. Hij beschrijft hoeveel krachtiger het web wordt als je er logica aan toevoegt. Het semantisch web is een uitbreiding van het huidige web, schrijft hij, waarbij computers de data van het net beter kunnen verwerken en 'begrijpen'.

“Weet je hoe een webpagina er voor een zoekmachine uitziet? Zo.” Frank van Harmelen laat een tekst in het Chinees zien. “Een computer heeft – net als jij – geen idee wat er staat. En voor hem zien alle webpagina’s er zo uit. Het enige waar hij goed in is, is woorden vergelijken: bijvoorbeeld of de zoekterm die jij intypt op een pagina voorkomt. Uiteindelijk rangschikt hij de gevonden websites op populariteit.” Wat zou het handig zijn wanneer de zoekmachine inzicht had in de tekst. “Stel dat de computer weet dat dit een tekst over malaria is. En dat er een label aan de eerste alinea hangt, dat hem vertelt dat daar de symptomen staan beschreven. Dan zou die zoekmachine jou, als je informatie over de symptomen van malaria zoekt, die alinea kunnen presenteren en niet alleen een link naar de hele pagina.” Een inhoudslabel aan een alinea is een voorbeeld van *metadata*: gegevens die inzicht geven in de inhoud van een tekst of beeld. In jargon heet zo'n label een *tag*. “Dankzij metadata krijgt de inhoud van het web betekenis voor de computer. Zo kan hij ook zelf verbanden leggen. Wij noemen dit *semantisch* of betekenisvol zoeken.” Van Harmelen is hoogleraar Kennisrepresentatie en redeneren en een van de VU'ers die zich bezighoudt met de ontwikkeling van het semantisch web – de slimme variant van het huidige internet.

2006 ***E-Culture wint prijs***

Op de internationale semantisch webconferentie wint het Nederlandse 'E-Culture' de eerste prijs. E-Culture is een demonstratieproject waarmee de collecties van vier kunstinstituten met semantische technieken te doorzoeken zijn. Hoogleraar Intelligente informatiesystemen Guus Schreiber van de VU bouwde deze toepassing samen met collega's van het Centrum voor Wiskunde en Informatica en de UvA.

“Toen wij met E-Culture die prijs wonnen, meesmuilden sommige mensen dat dit een voorbeeld was van een 'klassiek' semantisch websysteem. De theorieën wat er met een semantisch web kan, zijn al veel verder ontwikkeld. Maar aan alleen theorie heb je niets als er geen toepassingen worden gebouwd.” Guus Schreiber heeft grote plannen met E-Culture. Hij wil het project uitbouwen. De demo-versie geeft een indruk van wat semantisch zoeken is. Wie de zoekterm 'dark' invoert, krijgt net als bij Google een lijst van zoekresultaten. Maar dan wel netjes gerangschikt in categorieën, waaronder 'voorwerpen met 'dark' in de titel' en 'voorwerpen met donkere materialen'. Bijzonder is bovendien, dat E-Culture ook voorwerpen vindt, waarbij het woord 'dark' helemaal niet in de omschrijving voorkomt. Twee nachtelijke schilderijen van Vincent van Gogh bijvoorbeeld. De zoekmachine beschikt namelijk over een soort woordenboek, waarin staat aangegeven of begrippen aan elkaar gerelateerd zijn. Daardoor 'weet' de computer dat nachten donker zijn en geeft hij ook de nachtschilderijen als zoekresultaat.

2012 ***Tweehonderdste gebruiker aangesloten op cultuurnet***

E-Culture bereikt een nieuwe mijlpaal. Het Sprengel Museum uit Hannover sluit zich aan: dit is de tweehonderdste instelling die meewerkt aan het culturele netwerk. Het Europees Cultuurnet is een semantisch netwerk waarmee de collecties van culturele instellingen gemakkelijk digitaal te doorzoeken zijn. Ook het publiek heeft toegang gekregen tot de databases.

Semantische zoeksystemen beschikken over woordenboeken waarin staat hoe begrippen aan elkaar verwant zijn. Maar je kunt er ook



Internet is niet alleen een werkterrein voor informatici, het is een complete wereld

een echt woordenboek aan toevoegen: dan maakt het niet meer uit of je 'vrouw' of 'woman' als zoekterm opgeeft. “Het probleem met semantische technieken is de schaalbaarheid”, vertelt Schreiber. “Als je straks inderdaad tweehonderd instellingen hebt en acht talen, dan groeit de complexiteit.” Wetenschappers discussiëren of er ooit een semantisch systeem komt dat het hele web kan doorzoeken, zoals Google dat kan. Een project als E-Culture beslaat een overzichtelijk domein: in de strak geharkte omgeving van museumcollecties kan de semantische zoekmachine zijn weg wel vinden. Maar wat als de jungle van het web open staat? Dan komt de computer tegenstrijdigheden tegen: op de ene site staat dat kernenergie veilig is, terwijl andere sites het tegendeel beweren. Sommige sites zeggen dat Amsterdam de hoofdstad van Nederland is, maar andere sites gaan over Amsterdam in Texas. Zal een semantische zoekmachine dat verschil 'snappen'? Van Harmelen ziet de toekomst met vertrouwen tegemoet. “Nog geen tien jaar geleden werd een universele zoekmachine als Google ook voor onmogelijk gehouden. En waarom zou een zoekmachine niet semantisch kunnen zoeken? Wij zien toch zelf ook het verschil tussen Amsterdam in Noord-Holland en Amsterdam in Texas? Dat komt door de taal van de website, maar ook door begrippen op de pagina's. Zie je 'gracht' en 'Dam' in de tekst, dan gaat het waarschijnlijk over ons Amsterdam. Met onder andere statistische analysemethoden kunnen we computers dat trucje hopelijk ook leren.”



Meer weten?

U kunt het semantisch zoeken zelf uitproberen op e-culture.multimediam.nl, onder *Online demo*.

2013 **Autosites populairder door semantisch zoeken**

Nieuwe zoekmethoden hebben de verkoop van auto's via internet flink bevorderd. De afgelopen vijf jaar is de digitale verkoop met ruim een kwart toegenomen. Dat blijkt uit een onderzoek van de VU. Piet Akker, voorzitter van de Bovag, raadt internetkopers aan de auto te laten keuren: “Zorg dat u de koop kunt ontbinden als het resultaat tegenvalt.”

Van Harmelen en Schreiber verwachten dat het semantisch web als een eilandenrijk zal groeien. “De eerste eilandjes zijn er al. E-Culture richt zich op de cultuursector. Een bedrijf als British Telecom gebruikt semantische technologie voor het doorzoeken van hun intranet, dat reusachtig is. Zo zullen er meer eilanden ontstaan, binnen de wetenschap en binnen bedrijven.” Die eilanden zullen groeien: de vraag is of ze versmelten tot een groot semantisch rijk. Van Harmelen verwacht dat dat snel zou kunnen gaan wanneer de commercie de technologie adopteert. “Een bedrijf als Ilse wil semantische technologie gebruiken voor internetveilingen. Stel dat je op internet een auto wilt kopen, maar het type dat je zoekt is er niet. Als de site over een semantisch woordenboek over auto's beschikt, kan hij je vergelijkbare modellen van een ander merk laten zien.” De semantische technologie gaat al zo hard dat W3C, de organisatie die de spelregels voor het web vaststelt, werkt aan de standaarden voor metadata. Schreiber en Van Harmelen zijn hierbij betrokken.

2016 **Persbericht: VU opent Faculteit der Webwetenschappen**

De Vrije Universiteit Amsterdam opent op 1 september een nieuwe faculteit. De Faculteit der Webwetenschappen bundelt het onderzoek naar netwerken, zoals het internet.

Het onderzoek naar het semantisch web is nu ondergebracht bij informatica, maar Schreiber zou dat liever anders zien: “Internet is niet alleen een werkterrein voor informatici, het is een complete wereld. Het semantisch onderzoek doen we samen met taalkundigen en psychologen. Wij verwachten dat privacy een probleem wordt bij semantische technologie. Computersystemen zullen zich beter aanpassen aan de wensen en interesses van gebruikers, maar dat kan alleen als hij ook informatie over die gebruiker heeft. Het lijkt me prettig als er binnen de faculteit juristen zijn die weten hoe je met informatie op het web moet omgaan.” ◀