

Erik Bouwer

Web 3.0: intelligentie die vragen oproept

DE EERSTE ZOEKMACHINES DIE NIET KOMEN MET LOUTER VERWIJZINGEN NAAR WEBSITES MAAR MET CONCRETE ANTWOORDENSETS, ZIJN AL GEÏNTRODUCERD. STEEDS MEER GEGEVENSBRONNEN WORDEN GEKOPPELD.

ZIJN WE WERKELIJK OP WEG NAAR HET SEMANTISCH WEB? EN KUNNEN WE OVERZIEN WAT DE CONSEQUENTIES ZIJN?

De begrippen semantisch web en web 3.0 vormen een bron van misverstanden. Beide worden vaak gezien als een vorm van intelligent internet of een opstap in die richting. Die misverstanden maken ook meteen duidelijk waarom het semantisch web lastig is te realiseren. Intelligentie zal eerst gedefinieerd moeten worden: is dat een eigenschap exclusief voorbehouden aan levende organismen? Spelen emoties en motivaties daarbij een rol? En is kunstmatige intelligentie per definitie een beperkte vorm van intelligentie?

Een informatieomgeving die realtime is, steeds groter wordt en meer onderlinge verbanden krijgt, is per definitie rijker dan een statische dataset. Mobiele applicaties zullen daarom een grote rol spelen in het intelligente web, omdat daarmee steeds vaker data inclusief aanvullende informatie over tijd en plaats worden gegenereerd. Daarnaast wordt onderlinge verbondenheid van online datasets op twee manieren een belangrijk gegeven. Allereerst in de vorm van technologische infrastructuur – bijvoorbeeld door cloud computing – en anderzijds door dynamische datasets op inhoudelijke betekenis onderling te verbinden. Dit laatste gebied vormt de grootste uitdaging, want in het semantisch web komt het vooral aan op het kunnen leggen van betrouwbare betekenisvolle relaties tussen bronnen. Maar van echte intelligentie is geen sprake; eerder zou je kunnen zeggen dat het internet een soort collectieve intelligentie oplevert – met als kanttekening dat het web voor een groot deel ook gevuld is met onjuiste of waardeloze informatie.

Hoe dom is Google?

Wout Hofman is werkzaam bij TNO op het gebied van innovatie en interoperabiliteit: de samenwerking tussen mens en machine, tussen systemen en tussen organisaties. “Er zijn grenzen tussen organisaties, maar je zou het liefst alles zo open mogelijk willen houden. Bronnen van informatie zou je willen combineren om ze aan gebruikers aan te bie-

den. Die bronnen zijn vaak ongestructureerd. Met standaarden en metadata kom je een heel eind, maar dat is niet voldoende. Het semantisch web zou daarbij behulpzaam kunnen zijn.”

Om bronnen te koppelen moeten die bronnen wel dezelfde taal spreken of op zijn minst duidelijk maken waar de overeenkomsten en verschillen liggen. Metadata vormen niet meer dan de basis voor herkenning en vindbaarheid in een geïsoleerde set van gegevens. De intelligentie van een zoekmachine zoals Google is maar zeer beperkt: die is nog steeds gebaseerd op het kaartenbakidee, waarbij de intelligentie zich in een stelsel van slimme onderlinge verwijzingen bevindt. Voor veel statische begrippen zijn er al aparte en statische Wikipedia-pagina's gemaakt, maar hier leidt een vraag niet automatisch tot een antwoord op maat of zelfs tot het juiste antwoord (zowel in semantische zin als in het opzicht van betrouwbaarheid). Het semantisch web zou bij iedere denkbare vraag tot een Wikipedia-achtige pagina op maat moeten leiden. Een belangrijke eigenschap, want waar het web ons nu confronteert met duizenden verwijzingen zou het semantisch web in de respons op een vraag juist veel irrelevante informatie achterwege kunnen laten. Niet langer is een uitgebreid informatieaanbod het uitgangspunt, maar een beperking van dat aanbod. Zoekmachines zoals Wolfram Alpha zoeken op basis van een vraag informatie bijeen en presenteren deze in rapportvorm. Maar een internet dat werkelijk rekening houdt met de context is iets wat nog moet komen, stelt Hofman. “Denk aan het woord ‘Jaguar’, dat als zoekterm twee soorten resultaten kan opleveren: een technisch en een biologisch. Het huidige web werkt nog steeds op basis van woordherkenning. Als je dat wilt verbeteren, moet je semantische modellen gaan inbrengen.”

Wijsheid toevoegen

Omdat het intelligente web nog steeds een machine blijft, is er sprake van kunstmatige intelligentie: mensen moeten





deze intelligentie aanbrenge in een voor computers begrijpelijke vorm. Met andere woorden, web 3.0 is een fenomeen, dat alleen kan bestaan met door mensen aangebrachte verbanden en logica. Die verbanden kunnen op twee manieren tot stand komen: door gebruikers aangebracht tijdens het gebruik van webcontent en door ontwikkelaars aangebracht door datasets onderling met elkaar te verbinden en door ontologieën samen te stellen en op elkaar af te stemmen.

Wie gaat die wijsheid aan het web toevoegen? Dat is volgens Hofman allereerst een kwestie van hard werken: nieuwe modellen maken, zodat duidelijk wordt wat het begrip 'Jaguar' in biologisch en technisch opzicht inhoudt. "Semantische modellen zijn vaak voor een deel al aanwezig, bijvoorbeeld in de vorm van datamodellen in databases. In semantische modellen zijn echter ook de grenzen aangegeven van begrippen: de onderlinge samenhang en het onderscheid met andere begrippen uit andere modellen."

Definities kiezen

Social tagging, het aanbrenge van metadata door webgebruikers, zorgt niet voor de benodigde interoperabiliteit. Zelfs wanneer verschillende *communities* hun eigen

semantische modellen ontwikkelen, komt die interoperabiliteit niet tot leven; de kans is groot dat er dan meerdere 'werkelijkheden' blijven bestaan. Hoe lossen we dat op? Hofman: "Wat wil je bereiken? Je kunt werken met zelflerende *rankings*: denk aan de vragen die gesteld zijn en de meest gegeven antwoorden. Daar hoeft je geen metadata voor toe te voegen. Ik geloof het meest in een combinatie van dit soort ongestructureerde en meer gestructureerde benaderingen. Semantische modellen zouden door partijen moeten worden ontwikkeld die dezelfde belangen of interesses hebben. Het UWV kan bijvoorbeeld bij het begrip 'inkomen' een heel andere definitie hanteren dan de Belastingdienst. Definities en ook modellen kunnen elkaar overlappen, de kans is groot dat je kunstmatig links moet aanbrenge".

Partijen kunnen, zeker binnen de overheid, wel teruggrijpen op bestaande begrippenkaders. Als het gaat om het inkomen van een persoon, kan de overheid er bijvoorbeeld voor kiezen om bij 'persoon' de definities van de GBA als uitgangspunt te nemen.

Maar niet alle begrippen lenen zich voor ontologieën en logica. Hofman geeft als voorbeeld het concept 'in de buurt van'. Dit concept heeft bij 'de afstand tussen hotel en



Fotografie: Daniëlle van der Schans

strand' een andere set waarden dan bij 'de afstand tussen hotel en golfbaan'. In het eerste geval is vijf kilometer wellicht een bezwaar, in het tweede geval niet. Je kunt dit wel ondervangen door de gebruiker bij dergelijke begrippen naar een specificatie te vragen, al dan niet met gebruikmaking van Google Maps.

Ook bij een intelligenter web worden we dus gedeeltelijk gedwongen onze vragen beter te specificeren, of systemen komen bij ons terug met vragen waarvan de antwoorden tot uitsluitingen leiden, zodat het web ons verder de weg kan wijzen.

“Een andere mogelijkheid is dat een zoekmachine de zoekresultaten beter groepeerd en sorteert, zodat je als gebruiker bijvoorbeeld ziet dat er twee vormen van ‘Jaguar’ zijn, waarbij de zoekmachine voor beide vormen een zo volledig mogelijk beeld geeft”, stelt Hofman.

Maar verder ziet hij het bouwen van semantische modellen en ontologieën niet als iets wat burgers zouden moeten doen. Toch zijn er initiatieven op dit vlak, waarbij ontologieën op wiki-achtige wijze tot stand komen. Daarnaast zijn er bedrijven die voor allerlei onderwerpsgebieden ontologieën ontwikkelen en verkopen. Op beide vlakken wordt gebruikgemaakt van standaarden die worden onder-

Internet of things

Een bijkomende nieuwe eigenschap van het nieuwe internet is dat het niet alleen computers en mensen verbindt, maar een *internet of things* wordt. Zoals de pc langzaam maar zeker vervangen is door meerdere apparaten (waaronder de nadrukkelijk aanwezige smartphone), zullen ook andere entiteiten op het internet aanwezig zijn: datasets afkomstig van voorwerpen en machines of van andere mensen en organisaties.

Een aantal voorbeelden: een webapplicatie legt de hardlooperoute die je aflegt vast via je de mobiele telefoon en sensoren in je sportschoenen. Die gegevens zijn te combineren met adverteerdersinformatie (bestedingen in de supermarkt die op je route ligt of die je de volgende dag bezoekt); met gezondheidsinformatie (verander je leefpatroon en over een maand wordt je route verlengd); of met geografische data: hoe kun je vermijden dat je tijdens je rondje nat wordt door een regenbui; of wie van je bekenden zou je kunnen tegenkomen als je de geplande route tussentijds aanpast? Uiteraard blijft voor al deze vormen van connecties een gemeenschappelijke semantiek noodzakelijk.

steund door het W3C, waaronder OWL (Web Ontology Language, de standaard voor het opstellen van ontologieën).

Data, plaats en tijd

Een aspect dat ontbreekt in veel gegevenssets is de factor tijd. Wat gebeurt er met data in de loop van de tijd? Hofman geeft hierbij ‘gedateerde’ informatie, gepresenteerd door Google als voorbeeld: een vermelding over acteur Victor Reinier in relatie tot een voorval met zijn vriendin. De verwijzing naar deze informatie staat bovenaan in de zoekresultaten bij Google. Deze informatie kan heel belangrijk worden gevonden omdat ze vaak wordt geraadpleegd, maar het hoeft al lang niet meer actueel te zijn. Dit zoekresultaat houdt zichzelf in stand door de werking van Google. Nieuwe ontwikkelingen worden buitengesloten omdat er geen tijdsaspect wordt meegenomen en zelfs de webgebruiker voegt in dit voorbeeld door zijn gedrag ook niet bijster veel intelligentie toe.

Hofman ziet wel iets in het gebruik van tijdslijnen op het web, bijvoorbeeld in de vorm van metrokaarten inclusief *overstapstations* naar dwarsverbanden. Vreemd genoeg denken ook overheden in dit soort tijdslijnen, in de vorm van levensgebeurtenissen: je wordt geboren, gaat naar school, treedt in het huwelijk, gaat scheiden of komt te overlijden. Er zijn natuurlijk ook gebeurtenissen die complexer zijn: je gaat failliet, hebt een aanvullend inkomen

nodig en gaat scheiden tegelijk. “Dat zijn moeilijke situaties waar de overheid als instituut nog steeds lastig mee om kan gaan”, aldus Hofman, terwijl mensen die verschillende factoren prima kunnen combineren in een stappenplan of samenhangend antwoord.

Betrouwbaar?

Het web wordt met het toevoegen van inhoudelijke intelligentie alleen maar complexer. Met de wens alles ook inhoudelijk logisch aan elkaar te verbinden wordt bovendien de kans op fouten groter: als logische verbanden geautomatiseerd tot stand zijn gekomen en gebruikers daarop vertrouwen, is de kans aanwezig dat ze onjuiste informatie tot zich nemen. Zelf controleren wat op je beoogde vakantiebestemming de afstand tot het strand is via Google Maps is iets anders dan het semantisch web dat jouw interpretatie van ‘in de buurt van’ heeft toegepast. Hofman: “Als je alles gaat vastleggen wordt de kans op fouten groter. Dat heeft alles te maken met afstemming, en dus met communiceren”.

Die toenemende complexiteit door het koppelen van bronnen plaatst ook een ander fenomeen op de voorgrond: in hoeverre kunnen we nu al inschatten wat daar de consequenties van zijn? Het tempo waarmee besluitvorming over het gebruik van technologische mogelijkheden wordt genomen is veel hoger dan de mate waarin vooraf wordt geprobeerd de consequenties inzichtelijk te maken.

Op dit moment weten we als burger en internetgebruiker al nauwelijks meer hoeveel identiteiten en data we op het internet hebben geplaatst. Er staan documenten en foto’s online, je gebruikt sites als Slideshare, Flickr, je plaatst via je mobiel tijd- en plaatsgebonden berichten op Twitter, maar ook op fora en in nieuwsgroepen. Veel mensen hebben verschillende profielen zichtbaar op meerdere netwerksites. Dat zijn allemaal gegevens die ook door anderen (dus ook systemen) gebruikt kunnen worden en uit hun context kunnen worden gehaald, zonder dat de verschaffer van die gegevens er nog controle over heeft.

Anticiperen door analyseren

Wanneer je Twittert via je mobiel of gebruikmaakt van *location based services*, laat je steeds vaker je sporen na op het web, die verborgen informatie bevatten. Wat is over tien jaar de betekenis van tijd- en plaatsgebonden informatie? “Men weet alleen van jou wat je aan informatie vrijgeeft, maar veel mensen hebben daar inmiddels geen idee of beeld meer van”, zegt Hofman. Overheidsinformatie neemt daarbij een aparte positie in. “Online zou je je eigen databeleid of je eigen *privacy policy* moeten kunnen beheeren via een soort cockpit. Om erachter te komen wat de overheid over jou weet zou je inzage moeten hebben in die gegevens.”

Als we de factoren tijd en plaats steeds vaker toevoegen aan informatie (zoals de Belastingdienst, die met de vooraf ingevulde aangifte inkomstenbelasting een voorsprong in de tijd neemt op onze werkelijke situatie) is het waarschijnlijk dat niet alleen marketeers, maar ook de overheid ons gedrag gaat analyseren en erop zal anticiperen. “Analyses over ons gedrag worden steeds gemakkelijker door gegevens te koppelen. Denk aan rekeningrijden en de ov-chipkaart”, legt Hofman uit. De data zijn al beschikbaar en worden bovendien steeds vaker centraal vastgelegd en uitwisselbaar gemaakt; het is vooral de politiek die van moment tot moment kan bepalen welke combinaties en toepassingen wel en niet zijn geoorloofd. Het is de vraag in hoeverre we zicht hebben op de mogelijkheden en gevolgen van dat uitgangspunt. Hofman: “Die discussies worden wel gevoerd, maar niet zozeer in de maatschappij. Het beste moment daarvoor zal de maatschappij zelf moeten bepalen. Verder zal de maatschappij pas naderhand worden geconfronteerd met de bijeffecten”.

Semantisch web over tien jaar

Wat zijn de gevolgen voor informatieprofessionals? Moeten zij zich veel meer op het semantisch web richten? “Veel aandacht zal de komende tijd moeten uitgaan naar de evolutie van dat semantische web. Hoe ziet het er over tien jaar uit?”, zegt Hofman. Kenniswerkers zullen volgens hem wel steeds meer zelf kunnen doen; kennismanagers zullen daardoor een andere rol krijgen. Verdere ontsluiting van informatie – denk bijvoorbeeld aan de beschikbaarstelling van alle bibliotheek- en archiefinhoud – zou wel een verrijking kunnen zijn voor het semantisch web: het vergroot het aantal toepassingsmogelijkheden enorm. We kunnen straks bijvoorbeeld zien via Google Streetview hoe een stad er honderd jaar geleden uitzag. De daarvoor benodigde informatie is al beschikbaar.

Wat komt er na het semantisch web? Het is niet ondenkbaar dat we in 2040 het web benaderen op basis van gedachten. Met de juiste hardware en software kunnen we al organen en ledematen aansturen. Omgekeerd zou zo’n connectie – mits er tweewegverkeer mogelijk is – ertoe kunnen leiden dat je mensen kunt *hacken*. Een evolutie met kanttekeningen dus. **IK**