

INTERVIEW

'Het nieuwe zoeken het nieuwe vinden'

Guus Schreiber

Prof. Guus Schreiber en zijn mensen op de VU zorgen er voor dat het technische hart van het internet blijft kloppen.

Samen met vele internationale wetenschappers zorgen ze hoogleraar Web & Media aan de VU binnen W3C voor standaarden rondom html, xml, data-

beheer en -uitwisseling op het web. In de jaren negentig was hij als een van de eersten 'online', nu richt hij zich met zijn onderzoeksgroep op de volgende grote stap:

het nieuwe zoeken, ofwel web 3.0. Hij doet veel onderzoek met cultuurhistorische instellingen en zal ook een project starten met de bibliotheekwereld, zodra de fondsen zijn toegekend. //Collectie praat met hem over semantisch zoeken, web sciences, netneutraliteit en de toekomst van de bibliotheek in onze hectische online maatschappij.

'Technisch kunnen we het wel aan, maar menselijk niet'

door Herbert Boland

Bent u een lezer?

"Een fervente lezer, van huis uit. Maar dat is wat minder geworden door dat ik niet meer zo goed zie. Mijn favoriete genre is de biografie en dan met name de negentiende eeuw. Als je door middel van iemands leven kijkt naar de geschiedenis, ontdek je en begrijp je veel meer van het leven in een bepaalde periode."

Komt u nog wel eens in een bibliotheek?

"Nee, vroeger wel maar nu niet meer. Ik ben een groot liefhebber van opera en ik leende in de bibliotheek veel video's van voorstellingen."

Momenteel koop ik de boeken die ik wil lezen. Ik kan voor de boeken die ik zoek al nauwelijks meer terecht bij mijn favoriete boekwinkel. Als een boek een tijdje uit is, moeten ze het bestellen. Ik doe het nu via internet, ze zijn mij als klant kwijt. Kennelijk ben ik iemand die in 'the long tail' van het assortiment is terecht gekomen."

Waar houdt u zich in het wetenschappelijk werk momenteel mee bezig?

"Een van de belangrijkste zaken is het toegankelijk maken van cultureel erfgoed, of iets specifieker, collecties van musea. Daar zijn we nu al vele >



foto: Vincent van den Hoven

jaren op Europese schaal mee bezig, onder andere in het project *Europeana.eu*. Wat we willen bereiken is dat catalogi van cultuurinstellingen worden opengesteld en gekoppeld. Dat is nog best moeilijk, omdat iedere instelling verschillend met metadata omgaat. De volgende stap is om te komen tot kennisverrijking door middel van betekenisvol zoeken in al die informatie door het aan elkaar te koppelen. Waar we uiteindelijk naar toe willen is het plaatsen van de historische objecten in een samenhangend kader, ofwel de *narrative*, het verhaal.

Voor enkele Amsterdamse musea publiceren we de collectiedata op een voor het publiek toegankelijke manier en bouwen er applicaties op. Het Rijksmuseum is een van de gebruikers. Door middel van een demo-applicatie kun je een gepersonaliseerde rondleiding door het museum maken op basis van je voorkeuren.”

Welke plannen zijn er voor onderzoek in de bibliotheekwereld?

“We zijn bezig met een NWO-voorstel in samenwerking met SIOB, Bibliotheek.nl en het Huygens instituut. Het is nog wat vroeg om daar wat over te zeggen. Maar ik kan wel wat zeggen over wat we nu in de tv-wereld doen. Samen met RAI en BBC proberen we het zoeken in tv-programma’s te verbeteren. Wat je ziet is dat de rol van tv langzaam aan het verschuiven is. De *couch potato*, die alles kijkt wat hem wordt voorgeschoteld, sterft uit. Mijn kinderen kopen al geen tv meer. Op dit moment is tv alleen nog maar interessant voor live uitzendingen, daarna gaat ieder programma direct op Uitzending gemist. Maar als actieve consument en tv-kijker heb je hulp nodig om nog iets te kunnen vinden. We doen onderzoek naar een systematiek voor het zoeken in de broadcast-archieven, op dezelfde manier als bij collecties van musea.”

Hoe kom je tot de beste zoekuitkomsten?

“Je kunt uitkomsten van zoekopdrachten grofweg op drie manieren bepalen: statistisch, aanbevelingen van derden en op basis van onderlinge relaties van de content. Geen van de drie is zaligmakend, maar ze zijn wel goed complementair. Wat eruit komt biedt serendipiteit: er rolt meer uit dan je dacht en daardoor word je verrast en verrijkt. Begin te zoeken bij Gustav Mahler, vind vervolgens Kokoschka en Gropius dankzij hun muze Alma Mahler en zo ontvouwt de geschiedenis van die periode zich voor je. We noemen dat betekenisvol zoeken. Dat is iets wat je voor specifieke domeinen, bijvoorbeeld beeldende kunst, broadcast of literatuur, implementeert.”

Wat vormt de input voor het zoeken?

“Een bedrijf als Amazon komt op een weinig intelligente manier tot aanbevelingen. Ze kijken in je cookies wat je eerder hebt bekeken of wat je bij hen hebt gekocht. Verder bevelen ze aan wat ze je graag willen verkopen. De data van Amazon zijn niet bruikbaar op andere sites. Wij willen informatie van social media als Twitter, LinkedIn en Facebook koppelen, zodat we jouw interesseprofiel kunnen verrijken met dat van je vrienden. Op die manier kunnen wij op een gerichte manier naar

serendipiteit in de zoekuitkomsten streven. Dus zowel in de input- als aan de outputzijde van het zoeken streven naar het creëren van zoveel mogelijk betekenisvolle koppelingen. De laatste jaren zijn er web-protocollen ontwikkeld, zoals Oauth, waarin de toestemming voor het benaderen van elkaars data op social media wordt geregeld. Belangrijk daarbij is dat je zelf de controle hebt over welke gegevens gebruikt mogen worden.”

U pleit voor een onderzoeksgebied Web Science.

Wat moeten we ons daarbij voorstellen?

“Web Science, waar we in 2005 mee zijn gestart, is het bestuderen van het web vanuit meerdere disciplines. Dat is van belang omdat op inter-

net veel wetmatigheden uit de fysieke wereld niet opgaan. Economische, juridische en psychologische modellen moeten worden aangepast aan een wereld waarin geen fysieke grenzen bestaan. En vergeet de ethiek niet. Er is al veel onderzoek gedaan naar het gebrek aan normbesef dat ontstaat als mensen anoniem kunnen zijn.”

U noemt zichzelf webpurist. Op welke manier is het web ‘puur’?

“Essentieel is het begrip *netneutraliteit*, dat vormt de ruggengraat van het internet. Iedereen kan vrij informatie publiceren en iedereen heeft vrij toegang tot die informatie. Tim Berners-Lee heeft dat recentelijk in de ➤

Het Nieuwe Zoeken

Het Nieuwe Zoeken, web 3.0 oftewel het semantische web. We betreden het domein van Guus Schreiber, hoogleraar Web & Media aan de VU en zijn collega Frank Harmelen, beiden werkzaam in de Semantic Web Group. Met hun vijftig medewerkers vormen ze een van de leidende onderzoeksgroepen die wereldwijd op dit gebied actief zijn. Ze werken daarbij onder andere samen met mensen als Sir Tim Berners-Lee, de man die het World-Wide Web bedacht.

Bij het traditionele zoeken, wat we allemaal dagelijks een paar keer doen, tik je een zoekwoord en vervolgens komt de zoekmachine met een rij documenten of pagina's waarin de zoekwoorden voorkomen. Daarbij zijn ze ruwweg op populariteit geordend. Lui als we zijn bezoeken we meestal alleen de antwoorden op de eerste pagina. De populariteit van die antwoorden wordt daarmee nog weer eens verder versterkt. Er is een hele industrie ontstaan in het manipuleren van de zoekuitkomsten. Maar met die uitkomsten moet je wel zelf aan de slag om er betekenis aan te geven, Google doet dat slechts in beperkte mate voor je.

Bij betekenisvol zoeken krijg je het antwoord op je vraag niet als lange lijst, maar in verschillende, samenhangende categorieën. Zoek je op 'Picasso' dan krijg je de schilder, zijn werk, zijn muzen, vrienden, exposities en cultuur-historische betekenis. Misschien duikt er een tweede antwoord op dat alles vertelt over de auto met dezelfde naam. Of zelfs een derde, die alles vertelt over de steensoort picasso.

Meer krijgen dan je dacht, ofwel serendipiteit, dat is de essentie van het semantische web. Je zoekt iets, en geheel onverwacht krijg je andere bruikbare informatie. Dat is geen verdienste van je computer of van je browser. Dit moet achter de schermen, bij de opslag van de informatie zelf worden geregeld. HTML biedt daarvoor geen enkel houvast, XML slechts beperkt. De taal RDF is daarvoor bedacht, zie www.w3.org/RDF. Daarmee kan alle informatie worden gelabeld en onderling gerelateerd. Dat het geen toekomstmuziek is toont Facebook aan. Bij het klikken van de Like-knop wordt achter de schermen al RDF aangemaakt. Maar het zal niet snel gaan. Het coderen en verbinden van alle data die wereldwijd

beschikbaar zijn, kost zeer veel tijd. Vandaar dat het zal groeien vanuit eilanden, domeinen. Er zullen steeds meer domeinen ontstaan die groeien en onderling verbindingen aangaan: *linked data*.

Schreiber en zijn mensen werken momenteel aan het ordenen en coderen van collecties van Nederlandse musea. Daarmee wordt een schat aan culturele informatie voor iedereen beter toegankelijk en ontstaan nieuwe inzichten. Het Rijksmuseum is een van de gebruikers van semantisch zoeken. Je kunt via een demo-applicatie een gepersonaliseerde rondleiding krijgen.

Een ander interessant domein betreft overheidsdata. Burgers worden daarmee zelfstandiger en niet langer afhankelijk van ambtenaren: een echte kennisdemocratie.

Er ligt inmiddels ook een onderzoeksvoorstel voor de bibliotheekwereld, waarbij wordt samengewerkt met onder andere SIOB, Huijgens Instituut en Bibliotheek.nl. Het is in afwachting van NWO-subsidie.

INTERVIEW

Amerikaanse Senaat nog weer eens duidelijk gemaakt. Kijk eens hoe gevaarlijk Facebook bezig is. Ze bouwen aan een soort privaat internet binnen hun eigen muren. Als iedereen daarin meegaat, wordt het basis-idee van het web ondermijnd.”

En Google Books?

“Ze hebben een zeer groot aantal, veelal oudere boeken waar de rechten van zijn verlopen, ingescand. (Ook mijn boeken zijn er te vinden, zonder dat ik daar toestemming voor heb gegeven.) Op zich nobel, maar het nadeel daarvan is dat het alleen maar via de eigen Google-faciliteit is te raadplegen zodat ze er advertenties naast kunnen plaatsen. Ik vind dat ze daarin wel over de schreef gaat. Grote bibliotheken gaan daar in mee en geven de regie uit handen. Ik vind dat eigenlijk niet kunnen, want erfgoed is een publieke zaak. Maar dat is op wereldschaal helaas lastig te realiseren. De Europeana.eu portal probeert ook zoiets te doen. Ook voor hen is het lastig. Ze moeten namelijk alle Europese bloedgroepen gelijk behandelen. Hiervoor werd dit vooral gedaan door Archive.org, die ook een paar miljoen boeken heeft gedigitaliseerd. Voor bedrijven als Google is Archive.org dan ook een vervelende horzel. Maar we hebben dit soort horzels nodig. Amerikaanse NGO's opereren vanuit heel sterke principes. Ik ben wel eens jaloers dat we dit in Nederland en in het verdeelde Europa niet kunnen.”

Er is niemand de baas over het internet...

“Ja, en daarom moeten we misschien eerst doorslaan om te ontdekken wat we niet willen, ik denk aan de site Geenstijl. Aan de andere kant zie je heel veel mensen belangeloos tijd investeren om de kwaliteit voor andere mensen te verbeteren. De uiteindelijke optelsom zal positief zijn. Wij werken aan de technische standaarden, de ruggengraat van het web. Dat is volledig ‘in control’. Maar je hebt ook nog zoiets als de ethische standaarden. Wat ik begrijp van Jeroen van den Hoven (wetenschappelijk directeur van het 3TU Center for Ethics and Technology aan de

‘Gedrag van instant gratification van
de consument is beangstigend’

TU Delft, red.) is dat ontwikkeling van ethische waarden en regels tijd nodig heeft. Op internet zijn we nog volop bezig met het vaststellen van onze gezamenlijke waarden. Uit die waarden volgen dan de regels. Voorbeeld: wat is de waarde van persoonlijke vrijheid en wanneer tast je



de persoonlijke vrijheid aan in een open omgeving als het web? Daar zijn we nog lang niet uit. We kunnen het oplossen met wetgeving, zelfcorrectie of zelfs censuur. Er zijn webpuristen, waar ik ook toe behoor, die censuur het ergste vinden dat er is. Toch kan ik me voorstellen dat er extreme situaties zijn waarbij je er wel mee akkoord gaat.”

Heeft u zorgen over ontwikkelingen op het internet?

“Kijkend naar hoe bijvoorbeeld kranten en tijdschriften onder druk staan denk ik dat we heel snel moeten naar een goed systeem van micro



foto: Vincent van den Hoven

payments voor content. Mensen zouden bereid moeten zijn om te betalen voor informatie. Ik ben een voorstander voor het rekenen van een klein geldbedrag, één cent bijvoorbeeld, voor ieder verzonden e-mail. Je krijgt geen spam meer en je ontwikkelt een begrip van economische waarde op het web. Voor smsjes betaal je immers ook.”

**U bent een groot voorstander van de ‘kennisdemocratie’.
Wat houdt dat in?**

“Neem een voorbeeld uit de geneeskunde, wat ik ooit heb gestudeerd voor ik overstapte op IT. Het grote probleem in de spreekkamer van

een arts is dat mensen niet weten wat ze moeten vragen. Het is heel handig als je je van tevoren kunt inlezen. Dat is kennisdemocratie. Je hoeft niet alles te weten wat de ander weet, maar je kunt je wel verdiepen in zijn kennisraamwerk.

Je kunt een parallel trekken met de rol van de criticus in de boekenwereld, in relatie tot de groeiende rol van leesclubs en de online-communities. Je zult over een paar jaar een herwaardering zien van de professionele criticus. Hij heeft de achtergrondkennis en kent de geschiedenis, maar hij krijgt wel een andere rol. Of hij met die rol nog steeds een goede boterham kan verdienen is een andere vraag.”

Wat vindt u van de stelling van Carr, die in zijn boek *The Shallows* wil aantonen dat onze informatieverslaving ons concentratievermogen ondermijnt?

“Het is een interessante hypothese maar ik kan het niet uit eigen ervaring bevestigen. Er zijn ook mensen die het tegenoverstelde beweren: dat we steeds beter in staat zijn om te formuleren, plannen enzovoort. Ik vind dat mijn kennis, dankzij internet, de laatste jaren alleen maar is toegenomen. Het heeft ook mijn capaciteit tot redeneren en reflectie

‘We moeten snel naar een goed systeem van
micro payments voor content’

positief beïnvloed. Naarmate je meer weet, wordt het makkelijker om dingen te onthouden en makkelijker om dingen in perspectief te zien. Internet brengt ons zoveel aan kennis dat het juist een enorme stimulans is om diep na te denken.

Wat ik echter wel beangstigend vind, maar veel meer buiten internet dan daarbinnen, is het gedrag van *instant gratification* van de huidige consument. Mensen kunnen niet meer even wachten, het moet nu en onder hun voorwaarden. Ik merk het bij mijn eigen studenten. Het zijn onderwijsconsumenten geworden. Het moet prettig zijn, zo min mogelijk energie kosten en wanneer zij dat willen. Ik krijg tegenwoordig afmeldingen per e-mail, dat iemand bijvoorbeeld niet kan omdat hij een tennistoernooi moet spelen. Ik maak daar korte metten mee, maar ik maak me er niet populair mee.”

Wat zijn belangrijke ontwikkelingen voor de toekomst?

“De grote strijd zal gaan tussen de open wereld van bijvoorbeeld partijen als Google en de gesloten werelden, met Facebook als meest in het >

oog springende. Daarnaast verwacht ik dat er voor interessegebieden, denk aan geschiedenis, erfgoed, muziek of geneeskunde, binnen tien jaar speciale gebieds-Google's zullen komen. Google wordt op een gegeven moment te groot. Technisch kunnen we het wel aan, maar menselijk niet."

En wat is de toekomst voor de bibliotheek?

"Er zijn wel cultuurpessimisten die zeggen dat zodra de huidige generatie vijftig-plus uitsterft, boeken en klassieke muziek zullen verdwijnen. Maar ik geloof dat niet. We moeten naar een sociaal concept, een combi-

'Alles koppelen kan niet,
web 3.0 zal binnen domeinen ontstaan'

natie waar de traditionele bibliotheek met zoekfunctie en sociale functie samen komen. Je moet de fysieke bibliotheek en de online *presence* als

een geïntegreerd geheel beschouwen. Want mensen willen nog steeds graag bij elkaar komen. Daarmee blijft de fysieke plek behouden. Ik vind het heel belangrijk dat gemeenschappen op fietsafstand bibliotheek-faciliteiten hebben. Het is een verarming als die de komende jaren door bezuinigingen zouden verdwijnen. In de VS staat de *public library* op een voetstuk. Dat zou hier ook moeten zijn."//



In reactie op Interview met **Guus Schreiber**

Belangrijk onderzoek

Toen ik de uitspraken van Guus Schreiber had gelezen constateerde ik dat ik het met vrijwel alles wat hij zegt eens ben. Het werk dat Schreiber en anderen doen is van groot belang voor de online vindbaarheid en zichtbaarheid van de informatie die culturele organisaties hebben te bieden en zal mogelijk ook bijdragen aan de (deels) geautomatiseerde verspreiding van de expertise binnen die organisaties. Dat is een interessant gegeven, als je weet hoeveel moeite het bibliotheken, musea en archieven tot nu toe kost om in beeld te blijven bij doelgroepen op het web. Er

is een begin gemaakt met het werken met sociale media en onderzoek naar zoekmachineoptimalisatie, maar de online strijd om aandacht is meedogenloos en het handmatig selecteren van parels uit de vele voor Google verstopte databanken kost onevenredig veel tijd. Het op een eigentijdse manier delen van die selecties gaat ons nog niet zo goed af. Wat dat betreft kijk ik reikhalzend uit naar het genoemde onderzoeksvoorstel voor de bibliotheekwereld. Iets zegt me dat we alleen maar baat kunnen hebben bij dat voorstel.

Ook met Schreibers uitspraken over netneutraliteit,

microbetalingen en het informatiegedrag van de moderne consument ben ik het eens. Samengevat: Schreiber doet belangrijk werk en zegt veel zinnige dingen. Het zou goed zijn als de bibliotheekwereld, ook de openbare, nader kennis met hem maakt.

Edwin Mijnsbergen

Freelance informatiespecialist