

Huidige IT-systemen maken het mogelijk om zeer grote hoeveelheden gegevens op te slaan, te analyseren en te interpreteren. Dit wordt 'Big Data' genoemd. Professionals als artsen en beleidsmakers gebruiken deze berg gegevens om betere beslissingen te kunnen nemen en ook commerciële organisaties doen er hun voordeel mee. Maar ondertussen hebben overheid en burgers weinig grip op deze complexe gegevensstromen.

Door Anne Fleur van Veenstra

Big Brother is

Groei van de databerg

Supercomputer *Watson*, ontwikkeld door IBM, versloeg in 2011 twee topkandidaten in de Amerikaanse televisiequiz *Jeopardy!*. Net als in 1997, toen supercomputer *Deep Blue* (ook van IBM) grootmeester Garry Kasparov versloeg tijdens een schaaktoernooi, lukte het de machine om de mens te verslaan. *Deep Blue* kon destijds 200 miljoen schaakposities per seconde bepalen. *Watson* kan nog veel meer: de supercomputer kan hele lappen natuurlijke tekst analyseren en interpreteren en zo precieze antwoorden formuleren. Zo is bijvoorbeeld heel Wikipedia ingelezen.

Deze toename in capaciteit geldt niet alleen voor supercomputers, maar ook voor het wereldwijde dataverkeer. Dit is volgens IBM toegenomen van één exabyte (ongeveer een miljard gigabytes) per jaar in 2001, naar één per dag in 2013. Deze databerg groeit bijzonder snel. Volgens onderzoeksbureau IDC (*International Data Corporation*) heeft de wereld per twee dagen zo'n vijf miljard gigabyte extra opslagcapaciteit nodig. Vanwege deze snelle groei is 90% van de bestaande data minder dan twee jaar oud.

Big Data: mogelijkheden

De mogelijkheden van data-analyse van deze enorme gegevensberg, zoals het ontwikkelen van nieuwe diensten, worden ook als 'big' beschouwd. Die nieuwe diensten zijn er overigens al. Zo heeft Google het zogenaamde 'Pagerank algoritme' ontworpen, om websites op basis van hun populariteit te kunnen rangschikken. Facebook analyseert onze webgedrag en verkoopt vervolgens advertenties die inhaken op de interesses die daaruit af te leiden zijn. Websites die goedkope vliegtickets aanbieden, maken gebruik van de

datahoop door prijzen van verschillende luchtvaartmaatschappijen met elkaar vergelijken. En navigatiesystemen wijzen ons op basis van real-time en historische verkeersgegevens de kortste en snelste route.

Big Data gaat echter verder dan het slimmer weergeven van het bestaande informatieaanbod, zoals een zoekmachine doet: het combineren van verschillende informatiebronnen. Dit maakt het mogelijk om patronen te herkennen in gegevens die eerder niet zichtbaar waren. Of er wordt geautomatiseerd een precies antwoord op een



Beeld / Herman Wouters

watching you?

specifieke vraag geformuleerd. Een voorbeeld van zo'n gegevensanalyse is het detecteren van transacties die met gestolen credit cards zijn gedaan. Op basis van de omvang en de locatie van de transactie en op basis van de geschiedenis van de aankopen die iemand heeft gedaan, wordt bepaald of een transactie wel door de eigenaar van de credit card is gedaan.

Ook in de gezondheidszorg zijn de verwachtingen omtrent de mogelijkheden van de Big Data technologie hooggespannen. Er kan bijvoorbeeld ongestructureerde informatie uit foto's of scans gehaald worden, en grote hoeveelheden DNA-gegevens kunnen worden geanalyseerd. En er zijn interessante kansen in de financiële sector, waar, op basis van Big Data, betere investeringsbeslissingen kunnen worden genomen die een hoger rendement opleveren. In Nederland verwacht het Ministerie van BZK dat Big Data kan helpen een betere inschatting te maken van de effecten van beleid – zodat waar nodig nog tijdig kan worden bijgestuurd. Zelfs in de politiek heeft Big Data zijn intrede gedaan. Zo heeft President Obama zijn herverkiezing in 2012 mede te danken aan het ontwikkelen van specifieke software ('Narwhal' genaamd). Deze software verzamelde tijdens campagne-evenementen gegevens over stemgerechtigden en sponsors van de campagne die vervolgens *real-time* werd verwerkt. Hierdoor zetten de Democraten hun campagnemiddelen, zoals geld, reclamespotjes en vrijwilligers daar in, waar ze dachten dat ze meest zouden opleveren.

Big Brother is watching you

Big Data brengt echter ook de nodige risico's

met zich mee. Nu burgers massaal – bewust of onbewust – hun digitale gegevens op het net verspreiden en achterlaten, spelen privacyvraagstukken steeds vaker op. Veel nieuwe ontwikkelingen in sectoren als de gezondheidszorg (zoals elektronische patiëntendossiers), energiesector (slimme meters) en mobiliteit (rekeningrijden) zijn niet mogelijk zonder opslag en verwerking van persoonlijke gegevens. Daarom stelt de Europese Commissie op dit moment een nieuwe Richtlijn op voor de bescherming van persoonsgegevens.

Het linken van verschillende bronnen met persoonlijke gegevens kan zeer gedetailleerde en persoonlijke informatie over individuen opleveren. Wanneer status-updates van Facebook bijvoorbeeld gelinkt worden aan gegevens die verzameld worden door het gebruik van de OV-chipkaart, wordt het mogelijk om iemand digitaal 'te schaduwen'. Inmiddels bestaat ruim de helft van alle Big Data uit surveillancegegevens, vooral van bewakingscamera's. De veiligheidsindustrie en opsporingsdiensten kunnen Big Data gebruiken om zeer nauwkeurig mensen in de gaten te houden.

De vraag is echter in hoeverre nationale (of Europese) overheden nog invloed op deze ontwikkelingen hebben. Recent werd door onderzoekers van het Instituut voor Informatierecht (IvIR) van de Universiteit van Amsterdam gesteld dat de Amerikaanse overheid straks ook toegang kan krijgen tot persoonlijke gegevens wanneer deze in de Cloud zijn opgeslagen, zoals de medische gegevens in het Elektronisch Patiënten Dossier (EPD). Het EPD wordt gebouwd door het Amerikaanse bedrijf CSC, en daarmee valt het onder

'Vanwege de complexiteit van de algoritmes die ten grondslag liggen aan Big Data-toepassingen, hebben burgers minder inzicht in hoe de informatie die verschijnt eigenlijk is geselecteerd'

Amerikaanse wetgeving. Daarmee moet het zich ook aan de Patriot Act houden, wat betekent dat de Amerikaanse overheid bij vermoeden van terroristische activiteiten in principe bij die gegevens kan.

Het gevaar van de filterbubbel

Ook wanneer mensen zelf op grote schaal gegevens prijsgeven aan (meestal Amerikaanse) internetbedrijven, kunnen deze organisaties de informatie gebruiken voor commerciële doeleinden. Informatie kan bijvoorbeeld worden gebruikt voor prijsdifferentiatie, of voor het personaliseren van zoekresultaten. Een mogelijk tweede risico van Big Data is dan ook de ongelijkheid die ontstaat in het informatieaanbod: niet iedereen krijgt online dezelfde sites en gegevens te zien. Deze ongelijkheid wordt door internet-activist Eli Pariser in zijn gelijknamige boek de 'filterbubbel' genoemd. Deze ongelijkheid staat haaks op het idee van het internet als een vrijplaats voor het uitwisselen van informatie, zoals in de beginnende jaren van het net. Niet alleen in Iran ('halal web') en China ('the great firewall') censureert de overheid informatie. Internetonderzoeker Viktor Mayer-Schönberger van het Oxford Internet Institute laat zien dat Google en Twitter ook in Europese landen meewerken aan het blokkeren van informatie. Zo worden in Duitsland zoekresultaten waarin de Holocaust wordt ontkend (wat strafbaar is) geblokkeerd.

Google gebruikt Big Data om steeds betere zoekresultaten te kunnen tonen aan gebruikers. Zo heeft de zoekmachine onlangs een tweede algoritme aan haar diensten toegevoegd. Dit tweede algoritme maakt, in plaats van alleen een mathe-

'Big Data gaat verder dan het slimmer weergeven van het bestaande informatieaanbod. Door het combineren van bronnen is het mogelijk om patronen te herkennen in gegevens die eerder niet zichtbaar waren'

matische opsomming van zoekresultaten, ook inhoudelijke connecties tussen begrippen. Dergelijke zoekmechanismen zullen veel beter in staat zijn om mensen te helpen de juiste informatie te vinden in de steeds groter wordende databerg, maar versterken mogelijk ook de filterbubbel.

'Big' uitdagingen

Er zijn ten minste drie grote uitdagingen met betrekking tot Big Data. Ten eerste zullen overheden kritisch moeten kijken hoe deze ongrijpbare gegevensstromen toezicht op wetgeving beïnvloeden. Op dit moment draait de discussie vooral om online rechten, zoals in het debat rondom het herzien van de Europese privacywetgeving. Nationale overheden krijgen echter steeds minder invloed op het internet, dat per definitie internationaal is en daarmee de eigen juridische macht ontstijgt. Mede in dat kader zullen komende jaren belangrijke discussies gevoerd moeten worden over eigenaarschap, en toezicht daarop, van gegevens.

Ten tweede zullen consumenten zich meer bewust moeten worden van de gevolgen van het toenemende gebruik van Big Data voor henzelf. Enerzijds kan dit in hun voordeel uitpakken, zoals beter inzicht in de beste prijs via vergelijkingswebsites. Anderzijds kan deze trend ook nadelen hebben, zoals ongelijke toegang tot beschikbare informatie. Vanwege de complexiteit van de algoritmes die ten grondslag liggen aan Big Data-toepassingen, hebben burgers minder inzicht in hoe de informatie die verschijnt eigenlijk is geselecteerd (zie ook het artikel van Jan Otten). Informatie is steeds minder vaak het gevolg van onze eigen waarneming of zoekalgoritmes die

aansluiten bij onze preferenties, maar van systemen waar we geen grip op hebben en die bepaald worden door zakelijke belangen. Daarnaast zijn burgers zich dikwijls onbewust van de mogelijke gevaren van Big Data voor hun privacy. Op dit moment is de grootste schender van de privacy niet de overheid, maar de burger zelf die – bewust of onbewust – alles via sociale media deelt.

Tot slot zullen bedrijven en organisaties die grote hoeveelheden data verwerken ook moeten inzien wat de gevolgen voor hen zelf zijn. Big Data is geen toverformule die alleen maar voordelen brengt. Overheden, burgers en consumenten zullen steeds meer inzicht eisen in hoe informatie wordt gebruikt en wat er met hun persoonsgegevens gebeurt. Bedrijven zullen daarom transparanter moeten worden en duidelijk maken hoe zij omgaan met alle (persoons)gegevens. Zorgvuldige verwerking van gegevens zou bijvoorbeeld onderdeel kunnen worden van initiatieven van maatschappelijk verantwoord ondernemen. Voordat de voordelen van Big Data worden omarmd, zullen alle betrokkenen dus de nodige maatregelen moeten nemen.

Anne Fleur van Veenstra is onderzoeker en consultant bij TNO. Ze doet strategisch en beleidsonderzoek naar onderwerpen die betrekking hebben op de informatiemaatschappij, zoals privacy, e-government, open data en big data. Ze is gepromoveerd aan de Technische Universiteit Delft op onderzoek naar IT-geïnduceerde transformatie van de publieke sector.

Wat is Big Data?

Grote hoeveelheden digitale gegevens die opgeslagen worden en die extreem sterke processors nodig hebben om dit te verwerken, worden 'Big Data' genoemd. Het gaat niet alleen om informatie die bewust wordt gecreëerd (zoals filmpjes die geüpload worden), maar ook om gegevens die sensoren verzamelen, bijvoorbeeld over het weer, de verkeerssituatie, of GPS-signalen die mobiele

Literatuur

- Fischermann, T. & Hamann, G. (2012). "Neutrale Suchergebnisse sind eine Fiktion", *Die Zeit*, 13 september 2012.
- IBM Watson (2013). 'Ushering in a new era of computing', www-03.ibm.com/innovation/us/watson
- IBM (2013). 'What is big data?', 2013, www-01.ibm.com/software/data/bigdata
- IDC (2011). 'Extracting value from chaos', www.emc.com/collateral/demos/microsites/emc-digital-universe-2011/index.htm
- Hoboken, J.V.J. et al. (2012). 'Cloud computing in higher education and research institutes and the USA Patriot Act', Amsterdam: IvIR.
- Manyika, J. et al. (2011). 'Big Data: The next frontier for innovation, competition, and productivity', McKinsey Global Institute
- Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2012). 'Big Data: Wat moet de overheid ermee? Verslag van de rijksbrede expertbijeenkomst op 25 oktober 2012', B-16467.
- Hijink, M. (December 2012). 'Big Blue in Big Data', *NRC Handelsblad*.
- Pariser, E. (2011). 'The filter bubble. What the internet is hiding from you', New York: The Penguin Press,
- Madrigal, A. (2012). 'When the nerds go marching in', www.theatlantic.com/technology/archive/2012/11/when-the-nerds-go-marching-in/265325/#.UKt6LCu_6kU.email
- World Economic Forum (2011). 'Personal data: the emergence of a new asset class'.
- Xconomy, Roush, W. (2012). 'Google gets a second brain, changing everything about search', 12 december 2012, www.xconomy.com/san-francisco/2012/12/12/google-gets-a-second-brain-changing-everything-about-search

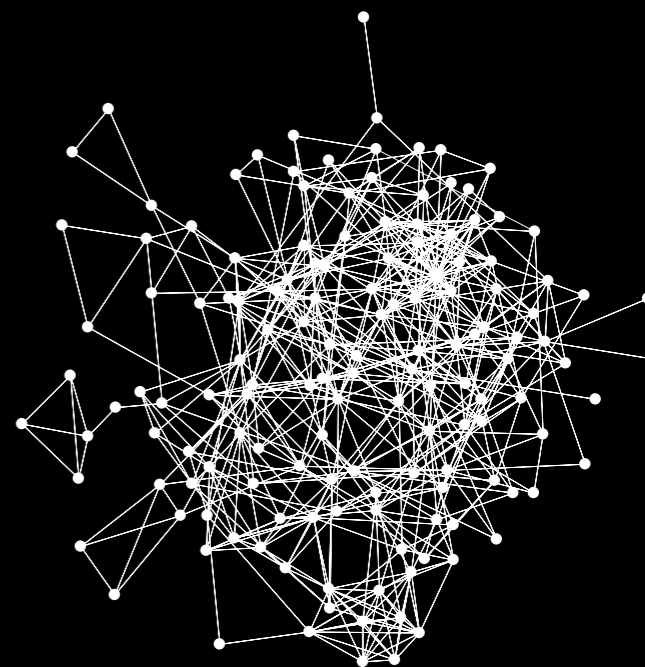
telefoons uitzenden. Verder heeft het *World Economic Forum* in 2010 uitgerekend dat er meer dan vijftig miljard apparaten aan het internet (het zogenaamde 'Internet of Things') zijn verbonden. De onbeperkte interactie tussen al die apparaten is een van de belangrijke oorzaken van de snel groeiende databrij.

Feiten & cijfers

Samenstelling: **Annemarie Bijloos**

Complexiteit van sociale netwerken

De complexiteit van onze belevingswereld wordt gevoed door de opkomst van *social media* als Facebook, LinkedIn en Twitter. Door deze media worden de sociale netwerken waarin individuen of groepen mensen met elkaar verbonden zijn, steeds groter en ruimtelijk verspreid. Waar mensen elkaar vroeger vooral ontmoetten op lokaal en fysiek nabij niveau, komen mensen nu ook op globaal en virtueel niveau met elkaar in contact. De manier waarop mensen zich tot elkaar verhouden, kan heel inzichtelijk worden weergegeven in zogeheten netwerkanalyses. Hierin wordt gewerkt met abstracte voorstellingen van sociale netwerken, ook wel grafen genoemd. Een graaf bestaat uit een verzameling punten, die met elkaar verbonden zijn door middel van lijnen. Een punt waar meerdere lijnen bijeen komen heet een knoop. Om inzicht te krijgen in de structuur van een bepaald netwerk, kan een graaf worden uitgetekend. Zo geeft onderstaande graaf het 'old boys' elite netwerk in het Nederlandse bedrijfsleven in 2007 weer: de punten zijn personen met minimaal 2 posities bij de top 250-bedrijven; de lijnen geven aan waar zij elkaar ontmoeten in raden van commissarissen of besturen. Een grafische weergave is mogelijk als het om grafen met tientallen of honderden knopen gaat. Moderne sociale netwerken worden



Figuur 1 Grafische weergave van een 'old boys' netwerk in het Nederlandse bedrijfsleven
Bron: M. Fennema & E. Heemskerker (2008), *Nieuwe netwerken: de elite en de ondergang van NV Nederland*