

ZOEKEN EN VINDEN | VINCENT M.A. JANSSEN

Terugblik op VOGIN-IP-lezing

'Ik zie een tweedeling in het programma van de VOGIN-IP-lezing 2018,' verkondigde Jacco van Ossenbruggen na afloop van zijn lezing over linked open data. 'Er is een scheiding tussen degenen die spreken over de kansen van nieuwe technologische ontwikkelingen en degenen die gevaren zien.' Hij zat er niet ver naast. De bezoekers werden op 9 maart jl. in de Openbare Bibliotheek Amsterdam zowel gealarmeerd als gerustgesteld over de trends in het informatieveld.

De zesde editie van de VOGIN-IP-lezing stond ook nu weer in het teken van 'zoeken en vinden'. De focus van het programma lag op de technologieën die het zoekproces mogelijk maken. Een thema waarmee keynotespreker Joseph Busch – consultant bij Taxonomy Strategies en 'bibliothecaris-by-heart' – het spits afbeet.

Grove instrumenten

Als 'the only English speaker on this conference' schudde Busch de nog slaperige (en supervolle) zaal wakker met een duidelijke visie op artificial intelligence en automatisering. Die twee begrippen zijn volgens hem niet hetzelfde en worden vaak door elkaar gehaald. 'Automatisering is datgene wat wij het meeste zien. Bij de meeste automatische processen draait het vooral om het herkennen van patronen. En bij slechts een klein gedeelte van deze automatisering komt kunstmatige intelligentie te pas.' Hoewel de keynotespreker de kracht van *cloud computing*, IBM Watson en het *Internet of Things* onderstreepte, zag hij meer mogelijkheden in het sa-

Foto: Eef Evers



Vincent M.A. Janssen
 Redacteur van IP en specialist
 Scientific Information bij de
 Maastricht University Library

Joseph Busch

menvoegen van reeds bekende automatiseringstechnieken: ‘*Keyword extraction* en sentimentanalyses zijn grove instrumenten. Maar als je ze samenvoegt, heb je krachtige tools die patronen kunnen herkennen.’

Patroonherkenning

Jacco van Ossenbruggen – die inviel voor zijn zieke collega Laura Hollink van het Centrum Wiskunde & Informatica (CWI) – zag zelf ook kansen voor het verbeteren van patroonherkenning. Hij deed dit echter liever met behulp van kunstmatige intelligentie. Kunstmatige intelligentie – dat volgens Van Ossenbruggen vroeger iets voor ‘losers’ was – wordt nu ingezet om een zogenaamde *concept drift* te analyseren. Hierbij gaat het erom de ontwikkeling van een concept door de tijd heen te bestuderen. Door het linken van open datasets hebben Van Ossenbruggen en zijn collega’s ontdekt dat het concept ‘wetenschap’ vroeger nauw verwant was aan religiestudies, terwijl het tegenwoordig juist geassocieerd wordt met seculier gedachtegoed. Door verschillende versies en betekenissen van een concept bijeen te brengen kunnen – met behulp van machine learning – zelfs voorspellingen gedaan worden over de toekomst van een begrip. Dit zou natuurlijk niet mogelijk zijn als de data nergens te vinden is. Maarten Dammers, vrijwilliger van Wikimedia, deed daarom in zijn lezing over Wikidata een oproep aan alle instellingen: ‘De kracht van Wikidata is de link tussen datasets. Publiceer daarom je data, zodat de Wikipedianen die kunnen vinden en verwerken.’

‘Filterbubbel bestaat niet’

In de praktijk leiden deze patroonherkenningen – vaak in de vorm van algoritmes – tot commotie en verontwaardiging. David Graus, datawetenschapper bij de FD Mediagroep, schoot evenwel in de verdediging. ‘De filterbubbel bestaat niet,’ zei hij stellig, en hij legde haarfijn uit waarom personalisatie van ons zoek- en vindgedrag juist noodzakelijk is. Zo zorgen de algoritmes – die onze online activiteiten monitoren – ervoor dat een zoekmachine toch met hoge precisie

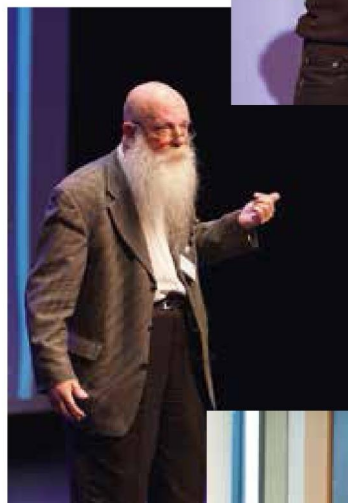


Foto: Eef Evers

Jacco van Ossenbruggen



David Graus



Arno Reuser



Christiaan Triebert



Elwin de Man en Rein Tellier

kan blijven werken, ondanks dat de hoeveelheid data toeneemt. Niet alleen voor de consument, maar ook voor de producent van informatie is personalisatie voordelig. Onderzoek toont bijvoorbeeld aan dat gepersonaliseerde nieuwsbrieven velen malen effectiever zijn.

Naast personalisatie kan ook automatische classificatie van teksten het zoeken en vinden verbeteren. Suzan Verberne, werkzaam bij het Leiden Institute of Advanced Computer Science (LIACS), liet in haar lezing daarom zien hoe dit classificatieproces werkt bij bijvoorbeeld spamdetectie, en hoe je hier zelf mee aan de slag kunt gaan. Eenvoudig is het niet, want je moet leren om te gaan met de programmeertaal Python.

Struikjes

Volgens Jacco van Ossenbruggen werd er tijdens de lezingen ook aandacht besteed aan de gevaren van dit soort gelinkte technologieën, zoals hacken en phishing. De aanwezigheid van Open Source Intelligence (OSINT)-goeroes Arno Reuser en Christiaan Triebert wekte inderdaad een dystopisch toekomstbeeld op; in hun workshops richtten ze zich op de bestrijding van (cyber)criminaliteit en terrorisme. Toch was hun boodschap die van een optimist: iedereen kan informatie met behulp van openbare bronnen verzamelen.

Christiaan Triebert van het internationale onderzoeksplatform Bellingcat liet in zijn workshop zien dat het niet zo lastig is om incidenten te analyseren met behulp van open geo-informatie. De deelnemers gingen zelf aan de slag met het lokaliseren van beeldmateriaal. Door te letten op details als struikjes, scheuren in muren en stickers op lantaarnpalen kon zomaar van menige vakantiefoto de exacte locatie bepaald worden.

Na deze oefening was het tijd voor het serieuze werk: het onderzoek naar MH17. Triebert liet de deelnemers in de voetsporen treden van de Bellingcat-onderzoekers. Met behulp van oude satellietbeelden, *reverse image searches*, RuTube, Google Earth en de stand van de zon zochten de deelnemers naar de route van de beruchte Buk-raket.

'Selectieve verantwoordiging'

Door te laten zien hoe eenvoudig OSINT kan zijn, spoorde Triebert zijn publiek aan om ook in de toekomst vaker gebruik te maken van dergelijke technieken. Maar helaas is nog niet iedereen zo bekend met de mogelijkheden van locatiegegevens. De onderzoeker bij Bellingcat liet zien hoe tracking-apps, zoals FitBit, zelfs de locaties verraden van geheime militaire eenheden. De zogeheten *heatmap* van Strava – een website die sportprestaties bijhoudt – toonde de joggingroutes van speciale eenheden in het Midden-Oosten. Dit was een gevaarlijke ontdekking, maar volgens Triebert absoluut niet de schuld van applicaties zoals Strava: 'Er wordt nog te weinig nagedacht over de consequenties van dit soort locatiegegevens. In plaats daarvan worden mensen nu boos op bijvoorbeeld Strava, terwijl deze website juist precies doet wat het belooft.' Hij noemt dit 'selectieve verantwoordiging'.

Evenals de speciale eenheden in het Midden-Oosten kampt ook onze eigen politie soms met een onzorgvuldigheid van haar personeel. Dat bleek uit het kijkje in de keuken bij de 'virtual agents' dat OSINT-agenten Rein Tellier en Elwin de Man gaven. Hoewel dienders heel voorzichtig dienen te manoeuvreren op het web, loopt het in de praktijk soms net even anders: 'Er zijn wel eens agenten geweest die hun eigen social media-accounts gebruikten bij online onderzoek. Toen wij hen vertelden wat voor gevolgen dit kon hebben, trokken ze wit weg.'

Infiltratie

Voorbeelden konden Tellier en De Man helaas niet geven; die waren geheim. Bij de aankondiging van hun lezing werd al medegedeeld dat er geen foto's van beide sprekers gemaakt mochten worden: 'Als we foto's op Twitter zien, krijg je geen sticker in je schriftje,' aldus dagvoorzitter Bart van der Meij. De twee 'virtual agents' lieten daarom zien hoe het interne proces van internetsurveillance werkt bij de politie. Wat blijkt: zij moeten aan heel veel juridische eisen voldoen. Zo mag een undercoveragent niet zomaar een verdachte volgen op Facebook; dat valt namelijk onder infiltratie.

Ook op technisch vlak worden hoge eisen gesteld. Bij de politie zijn ze continu



Maarten Dammers



Suzan Verberne



Bianca Kramer



Jeroen Bosman



Ruben Verborgh

op zoek naar de beste soft- en hardware-oplossingen waarmee ze niet alleen op kunnen gaan in de massa van gebruikers, maar ook de veiligheid van het personeel kunnen waarborgen. Welke systemen dat precies zijn, mochten de sprekers helaas niet vertellen. Ze gaven wel toe gebruik te maken van de tips en tricks van OSINT-specialist Arno Reuser, die op dat moment in een andere ruimte een workshop gaf over een slimme portable browser die gebruikt kan worden voor online onderzoek.

Zuckerberg-look

Allerlei trends en tools passeerden deze dag de revue. Ook ontkwamen de bezoekers niet aan een flinke portie blockchainjargon. Zo legden Bianca Kramer en Jeroen Bosman van de Universiteitsbibliotheek Utrecht in begrijpelijke taal de technologie achter deze nieuwe techniek uit. En voor wie deze workshop niet kon bijwonen, was er nog de tweede en laatste keynotespreker van de VOGIN-IP-lezing 2018.

Ruben Verborgh, professor semantisch-webtechnologie bij de Universiteit Gent, sloot de dag af met zijn visie over de blockchain. Verborgh, wiens nonchalante kledingstijl wel wat weghad van die van Mark Zuckerberg, liet zien hoe de blockchaintechnologie het evenwicht op het web kan herstellen. 'De balans tussen centrale spelers (zoals de techgiganten Google en Facebook) en de kleine nodes op het web is zoek. Hierdoor is een bijna gecentraliseerd web ontstaan,' aldus Verborgh. 'Maar als we de data van apps als Facebook weer in eigen handen krijgen, kunnen we het web herdecentraliseren. Hiervoor kan de blockchain ingezet worden.'

De bezoekers die Verborghs rappe betoog konden bijbenen, werden ten slotte getraakteerd op diens visie op digitale marktwerking. Het inzetten en linken van ongebruikte serverruimte zou volgens de Gentse professor informatie-uitwisseling naar nieuwe hoogtes kunnen tillen. Hij hoopt op het ontstaan van een wereldwijd 'web of data'. Maar voordat het zover is, moet er volgens Verborgh nog hard gewerkt worden aan 'de interoperabiliteit van data' en een betere samenwerking tussen de eigenaren van deze gegevens.

<