

# Zoeken en vinden:



Foto: Eef Evers

Annemarie van Campen

# 'Iedereen kan het'

Het thema 'zoeken en vinden' dekte volgens moderator Annemarie van Campen niet de lading van de VOGIN-IP-lezing op 21 maart jl. 'Er is dit jaar een werkwoord bij gekomen: checken.' Gelijk had ze. Het programma liet zien aan welke veranderingen de informatiesector onderhevig is.



**Vincent M.A. Janssen**  
Redacteur van IP en freelance  
informatieprofessional

**O**ok voor de zevende editie van de VOGIN-IP-lezing moesten de bezoekers in de Openbare Bibliotheek Amsterdam weer lastige keuzes maken. Werd het bijvoorbeeld een aantal lezingen over *data science* of toch een workshop zoekmachinemarketing? Ongeacht de keuze die de bezoeker maakte, hij werd door de keynotes in ieder geval getraakteerd op een portie factchecking en data(visualisatie).

### Personalisatie

Op het programma stonden deze keer een aantal heuse *data wizards*. Neem

Daan Odijk, *lead data scientist* bij RTL. Hij liet zien hoe bij zijn werkgever data en kunstmatige intelligentie worden ingezet om voorspellingen te doen over kijkersaantallen. Het weer, de seizoenen, feestdagen, voetbalwedstrijden; het is allemaal van invloed op het kijkgedrag. Verder gebruikt RTL data om zoveel mogelijk inzicht te krijgen in de behoeften van haar gebruikers. ‘Hiervoor hoef je maar te kijken naar de zoekresultaten van ons dochterbedrijf Videoland; die geven een schat aan informatie,’ legde Odijk uit. ‘Gebruikers zoeken bijvoorbeeld op *Game of Thrones*, een serie die Vi-

ontwaardigd. Maar ontkomen aan een gepersonaliseerd web lukt haast niet meer. Paula Kassenaar van de online fashion store Zalando liet in haar lezing zien hoe zij de webshop slimmer maakt met de nieuwste software: ‘Koop je een broek? Dan beveelt Zalando je meteen een bijpassend paar schoenen aan.’ Zo wordt de consument verleid om meer te gaan kopen.

### Ovengerechten

Data worden niet alleen gebruikt om meer omzet te genereren. Spreker Marieke van Erp (Digital Humanities Lab van de KNAW) maakte een data-

‘Uit acht miljoen krantenberichten kwamen maar liefst 27.000 recepten tevoorschijn’

werden, kwamen maar liefst 27.000 recepten tevoorschijn. De eerste analyses laten zien dat met name ovengerechten, kerstrecepten en vegetarisch eten het populairst waren.

De volgende stap is om deze data te verrijken en te koppelen aan andere informatiebronnen, bijvoorbeeld Wikipedia. Vooralsnog een lastige klus. In de digitale encyclopedie is namelijk weinig te vinden over etenswaren. Congresbezoeker Michel Wesseling, werkzaam voor Wikimedia, weet dit aan de scheve man-vrouwverhouding van bijdragers aan de online encyclopedie: ‘88 procent van de Wikipedia-



Daan Odijk



Paula Kassenaar



Marieke van Erp

deoland niet aanbiedt. Die data gaan dan vervolgens naar ons aankoopsteam om te kijken of ze de collectie nog beter kunnen laten aansluiten op de interesses van de gebruikers.’

De gebruikersdata worden ook ingezet om het aanbod te personaliseren. ‘Persoonlijke aanbevelingen leiden tot aantoonbaar meer views,’ weet Odijk. Toch zit niet iedereen op deze personalisatie te wachten. ‘Het schrikt mij enorm af om nu nog RTL of Videoland te gebruiken, als ik jouw verhaal hoor,’ reageerde een toeschouwer ver-

set van oude recepten. Waarom? ‘Wij [Marieke van Erp, Melvin Wevers en Hugo Huurdeman] zijn *foodies* en dol op eten’. Tijdens haar lezing liet zij zien hoe er de afgelopen decennia is geschreven over eten: ‘Hoe werd er in het verleden bijvoorbeeld gekeken naar gezond eten? Die informatie kan weer reuze interessant zijn voor dietisten,’ zegt Van Erp. Zij liet vijftig jaar aan recepten uit de Delpherkrantencollectie met de nieuwste taaltechnologie analyseren. Uit de acht miljoen krantenberichten die doorzocht

nen is man. Zij bepalen dus grotendeels wat encyclopediewaardig is. En dan krijg je vooral “mannenonderwerpen”,’ aldus Wesseling.

### Muzikale schatten

Naast de culinaire wereld was er tijdens deze VOGIN-IP-lezing ook aandacht voor muziek als onderwerp van data-onderzoek. Folgert Karsdorp (taalwetenschapper bij het Meertens Instituut) en Mike Kestemont (assistent professor in de literatuur bij Universiteit Antwerpen) ontwikkelden



een ‘deep learning-model’ dat rapteksten genereert. Het algoritme heeft hiervoor duizenden hiphopteksten geanalyseerd. Om aan te tonen hoe goed hun model werkt, namen zij de proef op de som op Lowlands. Festivalbezoekers ondergingen daar een soort Turingtest; zij moesten raden welke tekst geschreven was door een computer en welke door een mens. Slechts 61 procent van de antwoorden was correct. ‘Deelnemers deden het amper beter dan de gokkans,’ aldus de onderzoekers. Het VOGIN-IP-publiek deed het niet veel beter. Van de vier teksten die zij kregen voorgelegd, wis-

se dat niet, dan ben ik bang dat digitale muzikale schatten in de vergetelheid raken.’

### Datavisualisatie

Het communiceren van data is weer een heel andere tak van sport. Daar heb je bijvoorbeeld journalisten voor nodig. Een hiervan is keynote-spreker Frédéric Ruys, bekend van de VPRO-documentaireserie *Nederland van Boven*. Hij deed het werk van een datajournalist eenvoudig klinken: ‘Ik maak plaatjes.’ Maar daarvoor gebruikt hij wel gigantische en complexe datasets. Hij visualiseerde onder

van open bronnen-onderzoek te demonstreren. ‘Identificeren, checken, verspreiden; dat zijn de taken van het onderzoekscollectief Bellingcat,’ aldus Higgins. Tijdens zijn lezing liet hij bijvoorbeeld zien hoe zijn collega’s ‘*Netherlands’ most wanted*’ Shahin Gheybe – verdacht van een dubbele poging tot moord – wisten te lokaliseren aan de hand van de stand van de bloemen op een Instagramfoto. Higgins’ boodschap is inmiddels alom bekend: ‘Iedereen kan het’. Met een laptop, een wiferverbinding en engelen-geduld kun je verborgen dataschatten zoeken en vinden. <



Cynthia Liem



Mike Kestemont en Folgert Karsdorp



Frédéric Ruys

Foto's: Eef Evers

ten ze slechts in twee gevallen correct te raden door wie deze gemaakt was. Behalve rap stond ook klassieke muziek op het programma. Cynthia Liem, pianist en onderzoeker bij de TU Delft, maakt zich zorgen dat klassieke muziek slecht vindbaar is op het web. Zo zijn verschillende uitvoeringen van eenzelfde muziekstuk vaak niet van elkaar te onderscheiden in de metadata. In haar lezing liet zij zien hoe zij dit soort muziek verrijkt met bijvoorbeeld krantenberichten en beter vindbaar maakt in Google. ‘Doen

**‘Verdachte lokaliseren aan de hand van de stand van de bloemen op een Instagramfoto’**

andere potentiële dijkbreuken in Nederland, alsook een 3D-model van het Anne Frank Huis.

Naast datavisualisatie is er nog een vaardigheid die de journalist van vandaag dient te beheersen: OSINT. Oftewel Open Source Intelligence. Daarom werd deze dag afgesloten met een keynote van Bellingcat-oprichter Eliot Higgins. Zijn collega-factcheckers Henk van Ess en Christiaan Triebert stonden al eerder op het podium van de VOGIN-IP-lezing, dit jaar was het aan Higgins om de mogelijkheden

→ Wil je zelf ook een digitale detective worden? Dan kun je binnenkort in de leer bij Bellingcat. Zij hebben namelijk grootse ambities in Nederland. In een speciaal interview met Eliot Higgins vroeg IP naar de plannen van het platform in ons land. Zie volgende pagina.



Eliot Higgins

## Eliot Higgins:

# 'Ik merk dat Nederlanders me soms op straat herkennen'

Eliot Higgins, oprichter van onderzoeksplatform Bellingcat, lichte tijdens zijn keynote op de VOGIN-IP-lezing al eena tipje van de sluier op: Bellingcat opent een kantoor in ons land én gaat in een aantal Nederlandse steden workshops geven. IP vroeg hem naar zijn plannen.

Nederland heeft last van Bellingcat-koorts, mede dankzij het MH17-onderzoek en de recent verschenen documentaire *Bellingcat - Truth in a post-truth world*. En dat gaat niet onopgemerkt voorbij aan Higgins: 'Ik merk dat Nederlanders me nu soms op straat herkennen. Heel vreemd. Voorheen gebeurde dat alleen in Groot-Brittannië.' Hij zal hier aan moeten wennen, want het onderzoeksinitiatief gaat dit jaar uitbreiden naar Nederland.

### Voor de hand liggende keuze

Waarom Nederland? 'Toen we rond 2018 begonnen na te denken over een nieuw kantoor, kregen we vanuit Nederland al veel verzoeken om samen te werken. Die verzoeken kwamen van organisaties als universiteiten en nieuwszenders. Aangezien we al samenwerken met het Internationaal Strafhof in Den Haag, leek deze stad een voor de hand liggende keuze om ons kantoor te vestigen. In juni moet dit gaan gebeuren.'

'Nederland staat zeker in de top vier van bijdragers aan het platform'

De keuze voor Nederland had volgens Higgins niets te maken met een subsidie van de Nationale Postcode Loterij. 'Dat geld komt natuurlijk wel goed van pas bij de opening van ons nieuwe kantoor. Maar we krijgen ook geld uit andere hoeken, zoals van de Zweedse loterij. En ons onderzoek in Jemen blijkt ook heel populair te zijn onder sponsors.'

### Onvoldoende onderzoeksvaardigheden

De Nederlandse journalistiek kan – ondanks haar goede reputatie – nog veel leren van Bellingcat, meent Higgins. Ze gaan daarom jaarlijks tien vijfdaagse workshops organiseren. In deze trainingen wil Higgins verder kijken dan individuele onderzoeksvaardigheden. 'Als je als nieuwsorganisatie open bronnenonderzoek echt goed wilt uitvoeren, zijn de huidige vaardigheden van individuele journalisten niet voldoende. Je moet toegewijde onderzoeksteams oprichten. Ook op lokaal niveau. Daarom hopen wij nieuw leven te kunnen blazen in de lokale,

Nederlandse onderzoeksjournalistiek. Dat is een belangrijk doel van onze trainingen.'

### Trainers weggekaapt

De workshops die Bellingcat over de hele wereld geeft, zijn een succes. Maar dat heeft ook een keerzijde, weet Higgins. Nu redacties van bijvoorbeeld de BBC en de *New York Times* eigen 'mini-Bellingcat-afdelingen' oprichten, nemen zij de vrijwilligers van Bellingcat zelf in dienst. Dit zag Higgins bijvoorbeeld gebeuren met een van zijn bekendste trainers, Christiaan Triebert. Hij werkt sinds afgelopen maart voor de *New York Times*. 'Het roept bij mij gemengde gevoelens op. Het doel van Bellingcat is het opzetten van onderzoeksjournalistiek. En dat lukt goed. Maar het is jammer om mensen te zien gaan. We beschouwen hun nieuwe werkgevers niet als concurrenten, hè, want we zijn zelf geen mediabedrijf. We richten ons vooral op onderzoek, methodiek en training.'

### Goed op orde

Higgins merkt dat Nederland Bellingcat een warm hart toedraagt. Volgens hem staat ons land 'zeker in de top vier van bijdragers aan het platform'. Heeft hij al ideeën over het volgende land waar ze een kantoor zullen openen? Zover zijn ze voorlopig nog niet, zegt Higgins. 'Duitsland is een optie. Maar eerst moeten we ervoor zorgen dat we in Nederland alles goed op orde hebben.'