



'Ik geloof helemaal niks van niemand'

Eric Hennekam (58), archiefspecialist en docent 'zoeken op internet'.

„Vijftien jaar geleden stoorde ik me enorm aan een krantenartikel waar volgens mij niks van klopte. Ik ben niet zo'n ingezonden brievenrijver, ik wilde liever iets doen om de berichtgeving te verbeteren. Daarom bood ik bij de School voor Journalistiek in Utrecht aan om de studenten te leren hoe je goed in archieven kunt zoeken, ook online. Dat was toen nog nieuw en in een mum van tijd belden alle andere opleidingen journalistiek of ik daar ook wilde komen.

„Ik geloof helemaal niks van niemand, en zeker van journalisten niet. Ik help ze wel met plezier, geef vaak tips waar dingen gevonden kunnen worden en doe onderzoek voor veel media. Ook help ik regelmatig mensen met hun stamboom. Als ik eenmaal een archief gezien heb, dan vergeet ik het niet meer. Dat is wel een fijne afwijking. Ik werk vier dagen per week als medisch genealoog, dat is mijn echte baan. De rest van de tijd ben ik hiermee bezig.

„De belangrijkste vraag die je moet stellen is: welke personen zijn direct of indirect betrokken bij wat je zoekt? Als je parlementaire stukken nodig hebt, en je krijgt in Nederland nul op het rekest, probeer het dan eens bij de Europese Unie. Daar moeten vaak stukken naartoe gestuurd worden en worden ze mischien wel vrijgegeven. Of een ander voorbeeld: een tijdje geleden bleek dat sommige Chinese en Thaise massagesalons in Nederland en vooral zijn voor de *happy endings*. Een krant uit het Zuiden wilde weten of er daar ook zulke salons zijn. Je kunt dan gaan kijken naar bouwplannen en eigenaren vergelijken. Maar er zijn natuurlijk ook bezoekers bij betrokken. Toen heb ik mij ingeschreven op *hookers.nl* en jazerker, daar vond ik recensies over massagesalons. Er worden op veel meer plekken dan je denkt op de een of andere manier archieven bijgehouden.

„Je moet goed denken aan de historische context van wat je zoekt. Ik heb meegewerkt aan het onderzoek naar misbruik in de katholieke kerk. Veel journalisten gingen zoeken op het woord *internaat*. Maar in die tijd werden internaten meestal anders genoemd, dat was bijvoorbeeld *kostschool*, *juvenaat* of *klein-seminarie*. **Verplaatsen in de historische context geeft dus meer resultaat bij archiefzoek.** Goed archiefonderzoek begint met literatuuronderzoek.”

Laura Wismans

De fantasie van *Star Trek* en *The Matrix* komt dichterbij. Computers gaan u steeds beter begrijpen. Straks kunt u gewoon met zoekmachines praten.

Door Ewoud Sanders

Waar het mee zal eindigen weet niemand. Sommigen denken dat we uiteindelijk alle kennis van de wereld op een chip in onze hersens kunnen laden. Wil je een vechtsport of econometrie leren? Plug dan in op deze computer. *The Matrix*, maar dan echt.

Bij Google wordt al volop geëxperimenteerd met Google Glass, een brillmontuur met voor één oog een minischermje, vergezeld van een camera en microfoon. Gevraagd en ongevraagd projecteert die bril gegevens op het schermje: de buitentemperatuur, of je vaste treinverbinding is vertraagd, hoe je ergens naartoe kunt lopen, enzovoorts. Sergey Brin, een van de oprichters van Google, droeg het prototype al in april 2012 en naar verwachting komen de commerciële versies van deze futuristische bril eind 2013, begin 2014 op de markt.

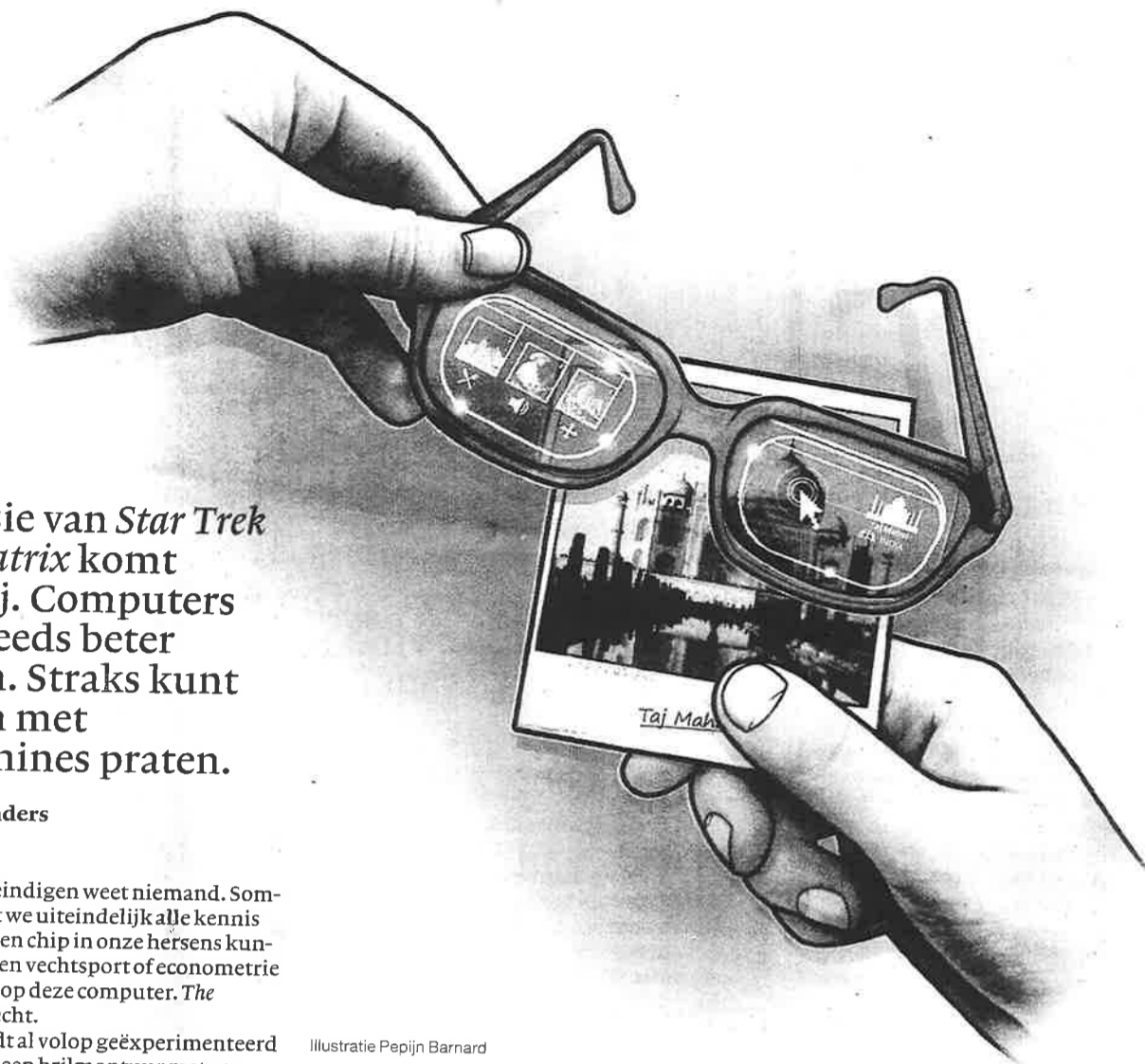
„Personalisatie is de toekomst van het zoeken op internet”, zei Marissa Mayer vorige week op het World Economic Forum in Zwitserland. In 1999 was Mayer de eerste vrouwelijke ingenieur bij Google, sinds vorig jaar is zij de hoogste baas bij concurrent Yahoo! Overigens maakt Yahoo! gebruik van de zoekmachine van Bing, van Microsoft. Alleen de presentatie en de omliggende informatie (afbeeldingen, video's, filtermogelijkheden) is van Yahoo!

De absolute marktleider onder de zoekmachines is Google. De percentages verschillen per land, maar 93 procent van de Nederlanders zoekt via Google op internet. „Ook wij zien personalisatie als de toekomst”, zegt Ben Gomes (44). Gomes is bij Google de rechterhand van Amit Singhal. Samen met een klein team bepalen zij de toekomststrategie van Google's zoekmachine.

Bij personalisatie maken zoekmachines niet alleen gebruik van het individuele zoek- en koopgedrag van een internetgebruiker, maar ook van gebruikers met een vergelijkbaar zoekprofiel. Je kunt dergelijke gepersonaliseerde zoekresultaten uitzetten – het makkelijkst bij Google – maar vaak zijn ze nuttig, vandaar dat zoekmachines er zo aan hechten.

Bij Google zetten ze nu hoog in op Google Knowledge Graph, een gigantische databank waarin ruim 570 miljoen 'entiteiten' (waaronder plaatsen, locaties, gebouwen, personen, tijdperken en kunstwerken) via ruim 18 miljard dwarsverbanden aan elkaar zijn gekoppeld. Zoek iets en rechts van de zoekresultaten krijg je ook informatie aangeboden uit Google Knowledge Graph: bouwstenen voor een volgende stap. Dit maakt het makkelijker om een onderscheid te maken tussen verschillende betekenissen die een zoekopdracht kan hebben. Zo kan Taj Mahal naar van alles verwijzen: naar het beroemde grafmonument in India, naar een bluesmuzikant, een casino, of talloze restaurants en hotels. De Amerikaanse, Franse, Italiaanse, Russische en Japanse versie van Google's zoekmachine geeft al resultaten uit de Knowledge Graph, in Nederland moet deze toepassing nog worden geïntroduceerd.

„We zijn er nog lang niet”, zegt Gomes opgewekt. „Maar dit is wel het begin. De hoeveelheid informatie in Google Knowledge Graph



Illustratie Pepijn Barnard

Een bril die alles weet

zal in een hoog tempo toenemen. Uiteindelijk willen wij zo exact mogelijk begrijpen wat iemand zoekt. Welke taak wil iemand op basis van de gepresenteerde informatie uitvoeren, in welke omstandigheden?”

In de trein

Bij zoekmachines denken veel mensen nu nog aan een toepassing op de laptop of pc. Maar steeds meer mensen zoeken mobiel, via hun smartphone of tablet. „De synchronisatie tussen mobiele en vaste apparaten zal dan ook een steeds grotere rol gaan spelen”, zegt Edgar Meij (31). Meij is nu nog als informaticus verbonden aan de Universiteit van Amsterdam, maar binnenkort gaat hij in Barcelona bij de researchafdeling van Yahoo! werken. „Op je werk zoek je iets op je pc”, verduidelijkt hij, „en in de trein naar huis kun je op je smartphone of tablet precies zien waar je gebleven was. En dus verder werken.”

Bij Google richten ze zich allang niet meer alleen op zoeken via pc's of laptops, zegt Gomes. Per drager of platform wordt er gewerkt aan nieuwe technieken. Zo zal spraakgestuurd zoeken nog sterk worden verbeterd. „Het is natuurlijk niet verstandig om een zoek-

opdracht in te tikken terwijl je een auto bestuurt”, grapt Gomes. Spraakgestuurd zoeken werkt al behoorlijk goed voor enkele wereldtalen, maar voor kleine talen als het Nederlands valt er nog een hoop te verbeteren.

„Het zal nog wel even duren voordat je met een computer kunt praten zoals in *Star Trek*”, voorziet Meij. „Maar ze zullen natuurlijke taal steeds beter gaan begrijpen. Je tikt dan vragen in zoals je ze in een gesprek zou stellen. Bijvoorbeeld: geef mij een overzicht van alle cafés die vrienden uit mijn geboortestad leuk vinden. Dat is wat Facebook nu probeert te bereiken met Graph Search.”

Wat er ook steeds meer zal komen, verwachten alle deskundigen, is zogenaamde *zero-query search*. Je loopt door een winkelstraat en je smartphone stuurt je deze boodschap: een tijdje geleden zocht je op internet naar een apparaat dat momenteel in een winkel 200 meter rechts van je in de aanbieding is. Andere toepassing: je loopt door Berlijn en je Facebook-app of Google+ meldt: vrienden van je hebben de volgende Berlijnse clubs en biercafés positief beoordeeld.

Zeker is dat het aantal grote spelers op de markt beperkt zal blijven. Google bestaat nog maar kort – sinds 1998 – en ook een andere zoekmachine zou dus in korte tijd megagroot kunnen worden, hoor je soms. Maar Google heeft inmiddels miljarden geïnvesteerd en heeft zoeken in die korte tijd breed geïntegreerd: je kunt zoeken door boeken, foto's, video's, wetenschappelijke publicaties, discussies, plattengronden, straten, satellietbeelden, enzovoorts.

Die enorme vaarsprong haalt niemand makkelijk meer in.

Smartphone en laptop in nabije toekomst veel beter gesynchroniseerd

Van Lycos tot Google

In 18 jaar tijd zijn zoekmachines geëvolueerd tot machines die begrijpen wat u bedoelt te vragen. Een korte geschiedenis.

Door Eric Sieverts

Sinds Google een bijna-monopoliepositie heeft verworven, lijkt het woord zoekmachine nauwelijks meer in het meervoud voor te komen. In de geschiedenis van het webzoeken is dat wel anders geweest, zoals de selectie oude zoekmachinelogo's hiernaast laat zien. Die geschiedenis begon in 1995 met Lycos, de eerste bruikbare zoekmachine voor het nog jonge web, gebouwd door een hoogleraar en studenten van Carnegie Mellon University in de VS. Onvoorstelbaar dat we toen opgewonden werden over een zoekmachine die 1,5 miljoen webpagina's doorzoekbaar maakte.

In de beginjaren speelde die grootte van zoekmachines een belangrijke rol. Wie had het grootste deel van het web weten te indexeren? Op de website van SearchEngineWatch (SEW) werd dat voor de belangrijkste webzoekmachines bijgehouden in staafdiagrammen en grafieken. Regelmatig namen ze daarin de koppositie van elkaar over. Die cijfers waren meestal afkomstig van de bedrijven zelf, die omvang als reclame beschouwden.

Als docent zoeken op internet aan de Hogeschool Amsterdam waarschuwde ik studenten in die beginjaren dat wat je aan informatie op internet kon vinden, veel minder was dan wat tegen betaling langs andere wegen online beschikbaar was. Maar die tijden waren snel voorbij. Met een verdubbeling elk jaar gaven webzoekmachines tien jaar later al toegang tot een veelvoud van wat betaalde bronnen boden.

Een verdubbeling aan doorzoekbare webpagina's elk jaar, waar komt dat cijfer vandaan? Aanvankelijk uit de SEW-gegevens. Maar na een tijdje werden geen groottes meer vermeld, omdat ze geen marketingargument meer vormden. De zoekmachines zetten toen in op relevantieordening: als het eerste resultaat het belangrijkste was, hinderde het niet als je minder had dan de concurrent. Nogal onzinnig, want als een zoekmachine het beste resultaat niet in zijn index heeft, heb je ook niets aan een sublieme relevantievolgorde.

Sorteren op relevantie

Webzoekmachines hebben een nieuwe manier van zoeken geïntroduceerd, doordat ze heel anders werkten dan de zoeksystemen waaraan we gewend waren. Er was geen ingewikkelde zoeksyntax meer nodig om een vraag te formuleren, zodat zoeken binnen ieders bereik kwam. En resultaten werden op hun vermoedelijke relevantie geordend.

Ook kregen gebruikers al gauw hulp bij het verbeteren van hun zoekvragen. Zoekmachines als Excite, InfoSeek, Euroferret en AltaVis-

ta (en later Teoma en Wisenut) kwamen na elke zoekvraag met rijtjes woorden die met de zoekvraag samenhangen. Die werden op basis van woordstatistiek berekend en waren meestal bedoeld om zoekresultaten in te perken, simpelweg door relevante woorden als aanvullende zoekwoorden aan te klikken. Na een paar jaar verdween die functionaliteit vaak weer geruisloos, of omdat die niet gebruikt werd, of omdat die zoekmachines verdwenen.

Na 2000 ging Google het zoeken steeds meer domineren, vooral door betere relevantieordening. Daarvoor werd niet alleen meer gebruikgemaakt van *waar en hoe* zoekwoorden in webpagina's voorkomen, maar ook van linkpatronen tussen webpagina's.

Grote zoekmachines – Google voorop – begonnen tijdens het intikken van een vraag ook suggesties te tonen voor precisering van de zoekvraag. Bij Google culmineert dat in het tonen van wellicht bedoelde resultaten terwijl de vraag nog wordt ingetikt – Google Instant. Daarnaast introduceerde Google technieken om de opbrengst van zoekvragen te verbeteren. Suggesties voor anders (en dus correcter?) gespelde zoekwoorden die meer opleveren. Automatisch toevoegen van enkel- en meervoud, woordvarianten, samengestelde woorden en geleidelijk ook synoniemen van de oorspronkelijke zoekwoorden. Wie op *vakantiehuis* zoekt vindt nu ook *chalet* en *bungalow*.

Automatisch zoeken op iets anders dan je als zoekvraag hebt ingerikt, omdat de zoekmachine denkt beter te weten wat je bedoelt, wordt door sommige zoekers verafschuwd. Wellicht om ook daaraan tegemoet te komen, introduceerde Google vorig jaar de optie om toch ook nog exact 'woord voor woord' te kunnen zoeken.

Zoekmachines beweren intussen eropuit te zijn antwoorden te geven, in plaats van lijstjes webpagina's te produceren. Bing afficheerde zich bij zijn introductie als een 'decision engine' en Wolfram|Alpha geeft echt alleen feiten. Niet eens echt nieuw, want al in 1996 streefde AskJeeves wat minder pretentieus datzelfde doel na. Ook Google zet intussen in op feitelijkheden met zijn Knowledge Graph. Op een vraag als *Beethoven*, krijg je naast een klassiek resultatenlijstje ook een blokje met feitelijke gegevens over de componist. Facebook gaat met zijn net geïntroduceerde Graph search ook dat pad op. Dat sluit allemaal aan bij de 'semantische' trend, waarbij semantische zoekmachines 'begrijpen' wat wij met onze zoekvragen bedoelen en ook wat in gevonden documenten wordt bedoeld. Om autonoom onze zoekopdrachten te kunnen interpreteren, willen zoeksystemen wel steeds meer van ons weten, of we dat willen of niet.

Eric Sieverts heeft ruim 30 jaar ervaring met het gebruiken, testen en beoordelen van zoeksystemen. Sinds 1994 geeft hij cursussen in het zoeken op internet.

ZOESPECIALIST



'Computers zoeken beter dan mensen'

Hans Henseler (48), partner bij cybersecuritybedrijf Fox-IT en lector e-discovery aan de Hogeschool van Amsterdam.

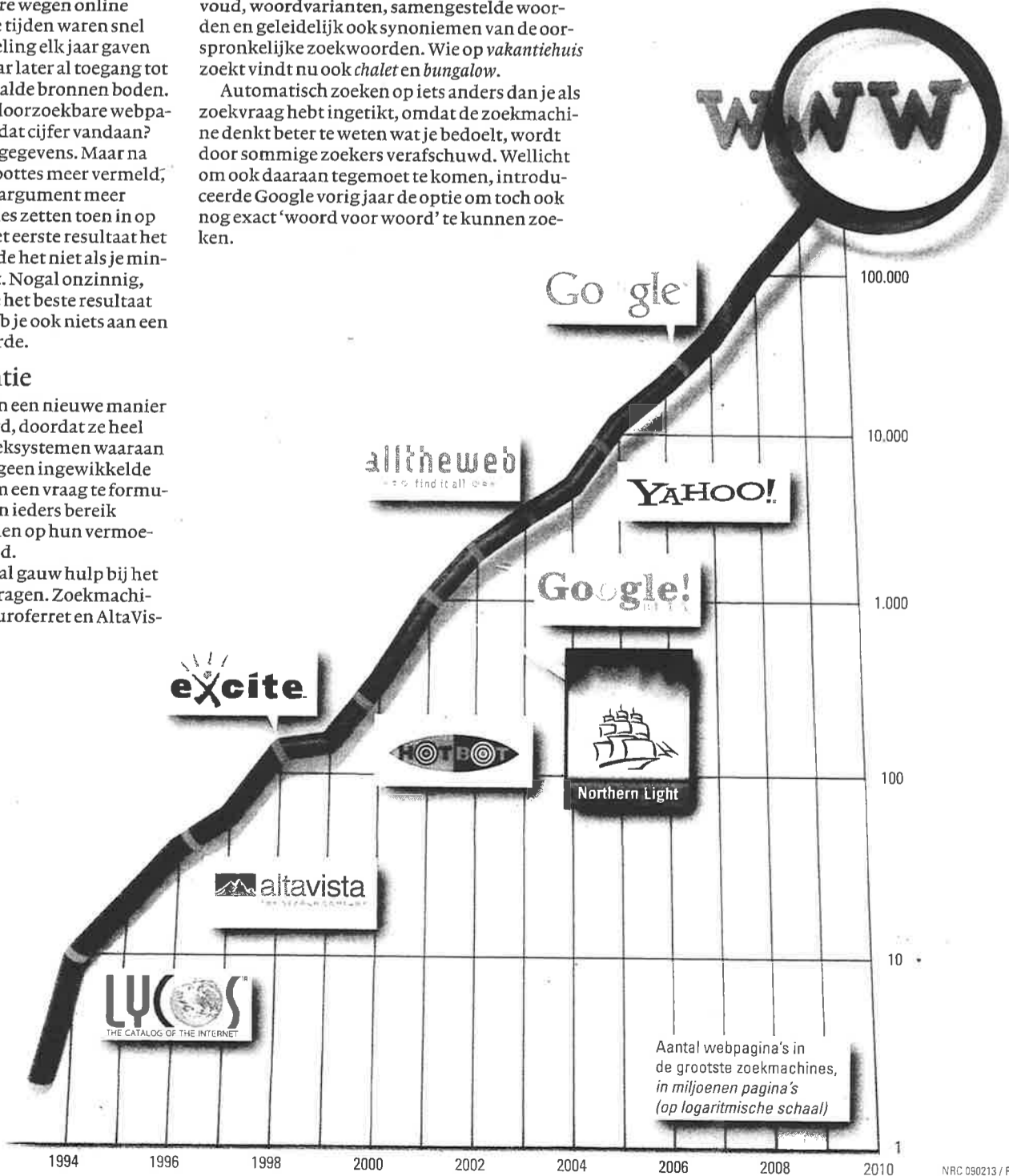
„Het laatste decennium is e-discovery enorm gegroeid. E-discovery is erop gericht om digitale informatie te vinden, te verzamelen en doorzoekbaar te maken, om het vervolgens te kunnen gebruiken als bewijsmateriaal. **Advocaten, forensisch accountants, bedrijven intern en natuurlijk de politie maken hier gebruik van.** Bij Fox-IT maken we software waarmee we e-discovery eenvoudiger maken. Op de Hogeschool van Amsterdam geef ik er les over aan studenten en organiseer ik in april voor het vierde jaar op rij een symposium. We zijn aan het kijken of we ook een post-hbo of postdoc-opleiding kunnen opzetten. Het is een populair onderwerp.

„Er is veel meer digitaal bewijsmateriaal voorhanden dan tien jaar geleden. Bedrijven maken gebruik van systemen om documenten op te slaan en te delen, **iedereen mailt en sms't zich suf.** Een smartphone is gewoon een kleine computer die je altijd bij je hebt, er wordt enorm veel op vastgelegd. Je moet je voorstellen: op een telefoon staat al gauw 8 gigabyte, dat is acht kasten vol met boeken.

„We koppelen een kabel aan een smartphone, pc of laptop en kopiëren alle gegevens, van whatsappberichten tot oude documenten. We halen ook zo veel mogelijk verwijderde bestanden terug. Met software wordt al die informatie doorzocht. **De politie of een deurwaarder kan bij een onderzoek of inval mensen verplichten alle telefoons en computers af te geven.** Advocaten hebben het recht die informatie ook te bekijken en te doorzoeken. Bij bedrijven die intern onderzoek willen doen naar bijvoorbeeld fraude geven mensen alles vaak vrijwillig af.

„De uitdaging is om uit het exponentieel groeiende woud van informatie net de interessante dingen te pikken. Je moet je bedenken, **de hoeveelheid digitale informatie verdubbelt ongeveer elke 18 maanden.** Het kan niet anders dan dat handmatig zoeken op termijn gaat verdwijnen. Er is gewoon te veel informatie. Daarom zijn we nu bezig om computerprogramma's te leren hoe ze zelf kunnen zoeken, en ook de context kunnen doorzien. Dat lukt steeds beter. En gelukkig maar want uit onderzoek is gebleken dat zonder hulp van slimme software mensen veel fouten maken en niet consequent zijn bij het beoordelen van grote bakken met informatie.”

Laura Wismans



ZOEKSPECIALIST



‘Verplaats je in wat je zoekt’.

Roelof Muis (40) doet als projectleider open source intelligence bij de Nationale Politie onderzoek via open bronnen als internet en sociale media.

„In 2011 begon dit project en sinds een jaar zijn we nu een officiële afdeling van 8 personen. Wij monitoren dreigingen op internet, gericht op personen, objecten en diensten. Dat doen we bij evenementen, zoals Koninginnedag of Prinsjesdag. Maar ook houden we constant in de gaten of er bijvoorbeeld dreigingen aan het adres van premier Rutte, of PVV-leider Wilders zijn. Hiervoor houden we onder meer de sociale media goed in het oog.

„We merkten al gauw dat het aantal berichten over bijvoorbeeld Wilders enorm groot was. Lang niet alles is een dreiging, maar alles moet wel doorgelezen en geduid worden. Wie uit het, heeft die persoon bijvoorbeeld een extreme culturele of religieuze opvatting? Vaak is het gewoon een grap. Software helpt ons om de stroom aan berichten behapbaar te houden en patronen en afwijkingen te signaleren. Hierin werken we samen met bedrijven, en onderzoekers bij de Radboud Universiteit en TNO. Zij helpen ons met linguïstische analyse – bijvoorbeeld combinaties van woorden waarop we alert moeten zijn. Als dergelijke woorden ergens geuit worden, dan worden wij automatisch geattendeerd. Daarna kijken we altijd zelf naar de context van zo’n uiting. De software wordt steeds slimmer en kan steeds meer automatisch.

„Regiokorpsen roepen daarnaast onze hulp in als ze bijvoorbeeld op zoek zijn naar vermiste personen, of na een diefstal. Als we zien dat een vermist persoon profielen op internet heeft, kijken we eerst of hij of zij nog online is en naar opvallende berichten in de timeline. Zo leren we veel en wordt de persoon vaak sneller teruggevonden. En bij diefstal zoeken wij bijvoorbeeld op Marktplaats naar de goederen door het merk, type en andere specifieke kenmerken in de gaten te houden. Soms krijgen wij weken na een diefstal ineens een signaal dat een gestolen auto te koop wordt aangeboden.

„Ik ben een zij-instromer bij de politie, ik kom uit de taalwetenschap. Ons werk is ook heel talig. Of je vindt wat je zoekt, hangt voor 80 procent af van hoe precies je de zoekopdracht formuleert. Verplaatsen in wat of wie je zoekt, is dus mijn belangrijkste zoektip. Gebruik je verstand.”

Laura Wismans

‘Denk als een document’

Véél informatie bij elkaar googelen kan iedereen, zegt zoekgoeroe Henk van Ess. „Het is de kunst om weinig te vinden, net dat wat relevant is.”

Door Laura Wismans

Collega’s en cursisten noemen hem ‘zoekgoeroe’. Henk van Ess (50) heeft na lang dubben ‘internetdeskundige’ op zijn visitekaartje gezet. „Ik heb een beroep dat ik zelf niet in één woord kan omschrijven. Ik kom langs als mensen zeggen: ‘Help ik verzuip in informatie.’” Media in binnen- en buitenland vragen hem hun journalisten beter te leren zoeken en vinden. Maar hij geeft ook cursussen aan universiteiten, artsen, bankiers en onderwijzers die beter willen leren zoeken op wat voor hun vakgebied handig is.

Waar komt uw internetdeskundigheid vandaan?

„Toen ik nog als journalist werkte – dertien jaar geleden – ging ik naar een persconferentie van een ziekenhuis dat claimde als eerste een operatie op internet te vertonen. Maar toen ik bij een zoekmachine die bewering intikte, kreeg ik 1.400 hits. Ziekenhuizen over de hele wereld beweerden hetzelfde, het was helemaal geen nieuws. Niemand op de persconferentie vroeg hoe dat zat. In die tijd was internet nog redelijk nieuw en ik besepte dat het een belangrijke rol kon spelen bij het verifiëren van stellingbeweringen. Neem niets klakkeloos aan, sinds die internetoperatie kun je die houding er bij mij niet meer uitslaan.”

Waarom wilt u anderen zo graag beter leren zoeken?

„Google laat maar een fractie zien van wat je kan vinden op het web. En binnen Google zelf moet je goed weten wat je doet. Ik kan moeilijk overweg met iemand die roept ‘Dat is niet zo!’, nadat hij een keer gegoogeld heeft. Zoekmachines zijn een soort kraan geworden. Die zet je open en daar stroomt dan informatie uit. Het is niet de kunst om veel te vinden. Het is de kunst om weinig te vinden, net dat wat relevant is. Dat moet iedereen weten, je doet jezelf tekort als je maar wat bij elkaar googelt. Het is onbegrijpelijk dat zoeken op internet op scholen, pabo’s en universiteiten geen verplicht vak is.”

Wat zou moeten worden onderwezen?

„Dat er een groot verschil zit tussen hoe een mens denkt en hoe een zoekmachine werkt. Mijn adagium is: denk als een document. Vaak zoek je vanuit jezelf, maar je moet juist zoeken vanuit de tekst waarnaar je op zoek bent.

„Dat leerde ik mijn zoon van 8 al, toen hij op school een soort Klokhuis-item moest maken. Hij had zich de vraag gesteld: hoe werken hersenen? Een ambitieuze vraag. Een van zijn vervolgvragen was: waarom raken hersenen vol? De Pavlov-reactie is het letterlijk intikken van die vraag in een zoekmachine. Dan kom je in een soort achterbuurt van het internet terecht, waar allerlei zelfbenoemde deskundigen wat roepen. In mijn cursussen stel ik altijd de vraag: welke zekerheid is er over het antwoord op een waarom-vraag? De enige zekerheid is dat het antwoord het woord ‘omdat’ of ‘doordat’ zal bevatten. Maar wie gebruikt er omdat in zijn zoekactie? Maar weinig mensen. Toch werkt het. Mijn zoon zou moeten zoeken op: ‘hersenen raken vol omdat’. Dan vind je tekst waarin uitleg staat. Vervolgens pas je filters toe om de betrouwbaarheid te verhogen.

„Wat ik dus aan wil leren, is nadenken over



Internetdeskundige Henk van Ess: ‘Google zelf ziet ook niet alles.’ Foto Bram Budel

het antwoord dat je wil vinden, en het taalgebruik van dat antwoord. Als je een constructietekening van een vliegtuig zoekt, tik je bij een zoekmachine al gauw ‘constructietekening vliegtuig’ in. Maar als je eens in een willekeurige constructietekening kijkt, dan zie je dat het woord ‘constructietekening’ zelf nooit op de tekening staat. Het is dus niet slim om op die term te zoeken. ‘Schaal’ en de schaal waarop je een tekening zoekt, is een betere zoekterm.

„En natuurlijk is een kritische houding belangrijk. Je moet niet snel iets aannemen. Op onze zoektocht naar hoe hersenen werken, drukten mijn zoon en ik op ‘bewerk’ in Wikipedia, de functie waarmee het lemma veranderd kan worden. Ik voegde toe dat er stoom uit je oren komt als je hersenen te vol raken. Mijn zoon schrok zich rot dat zo iets zomaar kan. Beseft dat iedereen een site kan beginnen en daar op kan zetten wat hij zelf wil.”

Verbeter de zoekkennis?

„Nee. Die wordt zelfs slechter. Bij mijn cursussen moet ik steeds meer dingen uitleggen die ik als bekend veronderstel. Collega’s in het buitenland ervaren hetzelfde. Het gemak waarmee informatie vindbaar is, maakt internetgebruikers waarschijnlijk gemakzuchtig. Je denkt vaak dat je vindt wat je zoekt, maar je bent je niet bewust van wat je niet vindt.”

‘Neem niets klakkeloos aan, dat is mijn houding’

Wat vinden we allemaal niet dan?

„Bij veel bedrijven is de zoekfunctie op de site een sluitpost waar weinig geld en aandacht aan wordt besteed. Dat zorgt voor problemen. Als ik bijvoorbeeld via de website van de Algemene Onderwijsbond probeer iets te weten te komen over salarissen in het onderwijs, ben ik blij met 250 hits. Maar als ik nu Google vraag wat deze ziet op hetzelfde woord en in dezelfde bron, krijg ik opeens zes keer zoveel gegevens. Het origineel – de AOB in dit geval – laat alleen de etalage zien, Google toont de hele winkel.

„Maar Google zelf ziet ook niet alles. De schatting is dat de Amerikaanse zoekmachine nog geen 50 procent van alle openbare bronnen ziet. De rest moet je van elders halen.”

Hoe vind je toch wat je zoekt?

„Door niet op onderwerp te zoeken, maar op het mogelijke antwoord. Als ik interviews zoek met Steve Jobs en de bron niet ken, ga ik naar Google. Je snapt dat ‘Jobs interviews’ ontspoort in links naar banensites. Maar ook een zorgvuldiger zoekactie als interview ‘Steve Jobs’ gaat mis. Waarom? Omdat in de meeste interviews niet het woord ‘interview’ staat. Je moet denken vanuit de kenmerken van het antwoord. En dan kom je op vondsten als ‘steve jobs’, ‘jobs says’ en omdat in elk interview de naam van de persoon, het woord ‘ik’, en dat wat hij zegt staat. Bekijk welke woorden bepalend zijn voor een willekeurig interview, tik die in.”

Hoe blijft u zelf op de hoogte van nieuwe zoekontwikkelingen?

„In een app op mijn iPad heb ik de namen ingevoerd van 500 specialisten op het gebied van zoeken en informatietechnologie die ik hoog aansla. Bij mijn ontbijt lees ik wat zij op internet gezegd hebben. De kracht van sociale media tegenwoordig is dat experts op hun beurt weer andere experts aanwijzen. Je krijgt zo een lijst fantastische bronnen.”