

Dokter Google en de strijd tegen corona

Brandende ogen, oorsuizingen, huiduitslag, schouderpijn ... Als je je niet lekker voelt, ga je tegenwoordig niet direct naar de dokter, maar zoek je eerst op het internet. Volgens enquêtes gebruikt twee derde van ons het internet om zich over mogelijke ziekten te informeren. Nu willen onderzoekers nagaan of "dokter Google" informatie kan geven over de verspreiding van het coronavirus of zelfs over voorheen onbekende symptomen. Wetenschappers in Harvard en Londen hebben in acht landen Google-zoekopdrachten onderzocht van mensen van wie is bevestigd dat ze symptomen hebben van Covid-19, zoals koorts, hoest of reukverlies. Het resultaat: het aantal zoekopdrachten naar Covid-19-symptomen is de afgelopen weken overal toegenomen. Die bevinding is in eerste instantie niet verrassend, omdat de angst om besmet te raken onder de bevolking algemeen is toegenomen en daarmee ook de noodzaak om symptomen te googelen. Interessant is echter wat de modellen van de onderzoekers zouden kunnen vertellen over de verspreiding en ontwikkeling van de pandemie. In een artikel in de *New York Times* beschouwde dataonderzoeker Seth Stephens-Davidowitz de studie als een kans om het zoeken naar symptomen in individuele Amerikaanse staten te linken aan de daar bevestigde Covid-19-gevallen. Hij ontdekte dat vragen naar reukverlies recent vaker voorkwamen in de Verenigde Staten, met name in de staten waar de meeste mensen besmet waren: New York, New Jersey en Louisiana.

Als men dat verband toepast op andere landen, schrijft Stephens-Davidowitz, zou men theoretisch twee dingen kunnen vaststellen: ten eerste of het aantal patiënten in een land of een regio hoger ligt dan voorheen bekend. Hij noemt Ecuador als voorbeeld: daar ligt het aantal bevestigde Covid-19-gevallen tien keer lager dan in Spanje, maar er wordt tien keer vaker gezocht naar geurverlies ("Ik kan niet ruiken"). Dat is een mogelijke aanwijzing dat veel mensen in Ecuador ziek zijn, maar nog niet getest werden. "De zoekgegevens suggereren dat Ecuador veel meer door Covid-19 getroffen zou kunnen zijn dan de officiële cijfers zeggen", schrijft Stephens-Davidowitz. Ten tweede zouden de zoekopdrachten symptomen kunnen onthullen die nog niet in verband zijn gebracht met de longziekte. In Italië bijvoorbeeld nam het aantal zoekopdrachten naar reukverlies toe voordat het als een bevestigd symptoom werd beschouwd. Stephens-Davidowitz ontdekte dat in de Amerikaanse corona-hotspots de afgelopen weken meer naar oogpijn gezocht werd. Dat was niet het geval in staten die minder getroffen zijn. Er zijn zelfs eerste rapporten dat oogslimvliesontsteking ook een symptoom van Covid-19 kan zijn. Om dat met zekerheid te kunnen zeggen, is de databank nog te klein. Over het algemeen waarschuwen zowel Seth Stephens-Davidowitz als de onderzoekers van Harvard en Londen ervoor om geen overhaaste conclusies te trekken uit Google-zoekopdrachten. Met reden.

De zoekmachine werd ooit gezien als een veelbelovend hulpmiddel om de verspreiding van een ziekte te analyseren, totdat men er uiteindelijk achter kwam dat de gegevens niet altijd even nauwkeurig waren. Zo lanceerde Google in 2008 Google Grieprends (*Google Flu Trends*). Google-ingenieurs wilden miljoenen zoekopdrachten gebruiken om te voorspellen wanneer en waar de volgende griepgolf zou uitbreken. De basis daarvoor werd gevormd door griepgerelateerde zoektermen. Het idee erachter: als veel mensen in een regio die termen plotseling googelen, zou daar binnenkort een griepgolf kunnen ontstaan. De bevolking, artsen

en autoriteiten kunnen er zich dan sneller en beter op voorbereiden. *Google Flu Trends* werkte in eerste instantie relatief goed; in veel gevallen vielen de voorspellingen samen met de latere analyses door de Amerikaanse centra voor ziektepreventie (*Centers for Disease Control and Prevention*). Al snel waren de voorspellingen echter steeds vaker verkeerd: sommige uitbraken werden helemaal niet herkend, terwijl *Google Flu Trends* andere, bijvoorbeeld in het griepseizoen 2012-2013, mateloos overdreef. In 2015 werd het project na aanhoudende kritiek beëindigd.

Zoals wetenschappers ontdekten, hadden de ontwikkelaars van *Google Flu Trends* niet gelet op twee dingen: aan de ene kant zorgde het aanbevelingsalgoritme van de zoekmachine voor onjuiste resultaten. Als je op zoek was naar "griepsymptomen", werd je aanbevolen om ook te zoeken naar "griepvaccinatie". Beide vragen werden in de analyse opgenomen, zodat het systeem zichzelf als het ware versterkte. Aan de andere kant waren er gewoon onjuiste correlaties tussen griep en andere seizoensgebonden zoektermen. " *Google Flu Trends* was deels griepdetector, deels winterdetector", schreven destijds critici zoals informaticus en politicoloog David Lazer. Is de huidige studie dan ook slechts een kijkje in de kristallen bol? Niet noodzakelijk. Ondertussen hebben onderzoekers meer ervaring met big-data-analyse en passen ze hun studies daaraan aan. In de in het begin genoemde studie gebruiken de wetenschappers niet alleen de absolute zoekopdrachten, maar harmoniseren ze deze ook met mediaberichten uit de respectieve landen op de dag van de zoekopdracht. Want hoe meer een onderwerp in de media aanwezig is, hoe meer ernaar gezocht wordt. Dat betekent dat er door correlaties onjuiste oorzakelijke verbanden kunnen ontstaan, zoals een voorbeeld uit Duitsland illustreert: op 16 maart verscheen in de *Frankfurter Allgemeine Zeitung* een interview met viroloog Hendrik Streeck, waarin hij zei dat hij nieuwe symptomen van Covid-19 had ontdekt, namelijk het verlies van geur en smaak. Wie de zoekopdrachten voor "geur-" en "smaakverlies" in *Google Flu Trends* vergelijkt, ziet een aanzienlijke toename vanaf 18 maart, twee dagen na het interview, toen ook andere media het rapport opgepikt hadden. Bijgevolg laten in dit geval de gegevens niet zien dat plotseling meer mensen in Duitsland Covid-19 hadden en hun symptomen googelden, maar alleen dat er een nieuwe ontwikkeling was die veel mensen interesseerde.

Het valt niet te ontkennen dat big-data-analyses kunnen helpen de verspreiding van het coronavirus beter te begrijpen. Berichten op sociale media en zoekmachines kunnen op nakende uitbraken wijzen, zoals eerste onderzoeken uit China aantonen. Artsen gebruiken al grote hoeveelheden data in combinatie met kunstmatige intelligentie om medicijnen en vaccins te vinden. En startups zoals BlueDot uit Canada doorzoeken dagelijks duizenden verschillende bronnen, nieuwssites, fora en blogs om erachter te komen of er zich ergens opvallende ontwikkelingen voordoen. Eind december konden ze een mogelijke uitbraak in Wuhan ontdekken. De critici van *Google Flu Trends* schreven in hun studie van 2014 ook dat de algoritmen van het project op zichzelf niet erg nauwkeurig waren. Maar als je ze combineert "met andere gezondheidsgegevens in real time", zijn de modellen zelfs beter dan die van de bevoegde instanties voor ziektepreventie destijds. Met andere woorden: zoeken naar symptomen in een hooiberg genaamd big data is niet helemaal nutteloos. De juiste technologie - en de juiste vragen - zijn belangrijk.