

OPMERKELIJK TIJDSCHRIFT VOOR BELANGWEKKENDE BIJZAKEN

Niet zomaar geloven wat Google belangrijk vindt

We zouden denken: een promovendus die Google wiskundig de oren wast over hun centrale algoritme, die hebben ze bij Google natuurlijk meteen in dienst genomen. Niks daarvan, zegt Joost Berkhout van de Vrije Universiteit. Vandaag verdedigt hij nog even zijn proefschrift, maar hij werkt in feite alweer een tijdje bij het Centrum voor Wiskunde en Informatica in Amsterdam. Aan de logistiek van veevoederfabrieken. Geen Google, toch leuk.

En eerlijk is eerlijk, aanvankelijk draaide zijn promotiewerk bij de sectie Bedrijfswetenschappen ook helemaal niet om wat de zoekgigant uit Mountain View uitspookt. Hij verdiepte zich in de wiskunde van Markov-ketens, een vorm van kansrekening die wordt gebruikt om netwerken te analyseren.

Daar was wiskundig nog wel het een en ander te verhapstukken, maar Berkhout ontdekte al lezend dat hij ook in de hartstreek van Googles zoektechniek was beland. Het centrale algoritme waarop Larry Page en Sergei Brin van Stanford in 1998 hun Google-zoekmachine baseerden, gebruikte ook Markov-ketens om zoekresultaten te ordenen van belangrijk naar onbelangrijk. Die aanpak was een revolutie in het internet-zoeken. Voorheen telden zoekmachines domweg het aantal weblinks op pagina's, waarmee slimme webmanagers driftig vals speelden. Brin en Page bepaalden ook hoe belangrijk die links waren.

Bij de Page-Brin-methode, legt Berkhout uit, wordt een denkbeeldige web-surfer losgelaten die vanuit een webpagina op een van de hyperlinks klikt en naar een nieuwe pagina springt, daar weer op een link klikt, enzovoorts. Pagi-

na's waar de surfer veel langskomt, zijn belangrijker voor een zoektocht dan stilere pagina's, luidt het basisidee. Die zijn het bezoeken waard en leveren ook veel internetverkeer op.

Klopt, concludeert de VU-promovendus, maar het gaat niet altijd goed. In elk netwerk bestaan eilanden van samenhangende webpagina's, waarnaar andere pagina's niet verwijzen, maar die zelf wel verwijzen. Een Googlesurfer die daar toevallig begint, komt er nooit meer terug. 'Zulke domeinen komen bij een Googlesearch niet snel bovenaan de ranglijst', zegt Berkhout.

Moeten we het centrale algoritme van Google, PageRank genaamd (naar oprichter Larry en niet naar het Engelse woord voor pagina), dus wantrouwen? En dan niet omdat er misschien ook commerciële redenen zijn om sommige resultaten hoger op in de lijst te zetten, zoals Googlecritici vermoeden?

Berkhout: 'Wat Google precies doet, is bedrijfsgeheim. Ik heb het over het originele algoritme, dat nog wel gepubliceerd is. Je moet aannemen dat er wel veel aan toegevoegd en gesleuteld is. Maar de kern van PageRank is er nog, dat is hun kapitaal.'

Het gaat de VU-promovendus niet om Google, maar om de wetenschapswereld waar PageRank algemeen gebruikt wordt zonder dat iemand beseft dat het ook dingen over het hoofd ziet. Berkhout ontwikkelde een alternatief, dat bij een zoekopdracht ook eenzame informatie-eilanden aanbeveelt als het nodig is. Voor de liefhebbers, hijzelf is voorlopig druk. Met de logistiek van het veevoer dus.

Martijn van Calmthout

UIT DE VERGETELHEID

Jaarlijks publiceren Nederlandse wetenschappers 172 duizend onderzoeken. In deze rubriek een greep uit de ontdekkingen die bijna onopgemerkt waren gebleven.

Wie?

Joost Berkhout (26)

Wat is zijn specialiteit?

Operations research

Originele titel publicatie?

'Topics in Markov Chain Theory and Simulation Optimisation' (proefschrift VU)

Vrij vertaald?

Waarom Google dingen over het hoofd ziet

