

Open toegang tot data in China?

In de laatste week van oktober kwamen in Peking zo'n zeshonderd academici uit alle werelddelen bijeen op de zogenaamde CODATA-conferentie over vrije toegang tot wetenschappelijke informatie. Ik was daar één van en ik was vooral nieuwsgierig naar de beschikbaarheid van data en andere informatie achter the Great Chinese Firewall.

Ik kan me voorstellen dat China alle belang heeft bij een zo open mogelijke toegang tot wetenschappelijke data. Van westerse onderzoekers want de toegang tot Chinese data voor het Westen wordt natuurlijk in de eerste plaats beperkt door de taalbarrière. Maar als het gaat om maatschappelijke of politieke informatie is de reputatie van China bepaald niet onbesproken.

Zo blijkt uit onderzoek van Zittrain en Edelman in 2002 dat zo'n twintig duizend websites niet vanuit China bereikbaar waren maar wel vanuit de

In een landenstudie van het *Open-Net Initiative* (ONI) staat dat China's Internet filtering regime is the most sophisticated effort of its kind in the world. Er zijn tal van organisaties en duizenden medewerkers op verschillende niveaus bij betrokken. Het filterregime schijnt heel fijnmazig te zijn, over de tijd steeds anders te worden ingesteld en zowel plaats te vinden op het niveau van het backbone van het Chinese netwerk als op dat van internet service providers. Er wordt ook op zoektermen gefilterd en Chinese internetcafés zijn wettelijk verplicht om het zoekgedrag van hun klanten zestig dagen te bewaren.


Maar in hoeverre is webinformatie nu eigenlijk effectief af te schermen? Van spamfilters weten we dat ze gebrekkig werken. En hetzelfde geldt voor de filters die angstige ouders installeren om hun kinderen weg te houden van smoezelige websites: kinderen

het voortbestaan van het Rijk van het Midden. Maar ook 'Taiwanese independence', Falun Gong en Dalai Lama leverden talloze hits op, al weet je bij Google natuurlijk nooit wat je *niet* vindt. Mijn experiment was niet wetenschappelijk in de zin dat elders op de aardbol iemand gelijktijdig dezelfde zoektermen intypte.

En de Chinese Google? Dat werkt immers schaamteloos mee aan de Chinese censuur? Bij de zoekmachine voor afbeeldingen valt het verschil duidelijk te zien. De Chinese Google levert slechts 173 plaatjes op voor het trefwoord 'Falun Gong', vooral van aanhangers van de sekte die moordaanslagen hebben gepleegd; de Nederlandse Google levert 45.200 afbeeldingen.

Bij het zoeken naar tekst verandert het beeld. Als je geen Chinees kunt lezen is het natuurlijk moeilijk na te gaan wat je krijgt, maar het intypen van Engelse termen geeft telkens vergelijkbare aantallen hits in de Chinese en de Nederlandse Googles. Zou het verfijnde systeem tijdens de CODATA-conferentie even op een laag pitje zijn gezet, tenminste in het hotel waar de conferentie werd gehouden? Of kon men zelfs per kamer op een knopje drukken: in 1707 zit een westerling, die hoeft niet te worden geblokkeerd? *Reporters sans frontières* meldde op 6 juni dat Google.com sinds 31 mei werd geblokkeerd in verband met de herdenking van het bloedbad op het Tiananmenplein op 4 juni, zeventien jaar geleden. Maar een paar dagen na de herdenkingsdag was deze blokkade opgeheven, blijkbaar omdat de urgentie van de censurering toen alweer was weggeëbd.

De CODATA-conferentie maakte duidelijk dat China vrije toegang tot wetenschappelijke data nastreeft. Zo ver is het nog niet voor alle soorten informatie. Het is mogelijk om de web-blokkades te omzeilen. Maar het is natuurlijk wel de vraag of de gemiddelde Chinese internetter dat aandurft. (Peter Doorn)

 www.codataweb.org/o6conf/
<http://cyber.law.harvard.edu/filtering/china/>
www.opennetinitiative.net/studies/china/
www.rsf.org/article.php3?id_article=17936



VS: vooral websites met nieuws, politiek, religie, gezondheid, handel en entertainment waaronder sex. De sites van Amnesty International, het Centre for Anti-Communism, de Taiwan Student Club in Austria en de Voice of America hoorden daarbij. De Chinese overheid voert een politiek van blokkades, die echter vooral effectief zijn tegen gebruikers die ze niet proberen te omzeilen. De blokkeringsystemen worden ook steeds verfijnder al vergen ze veel onderhoud, waaronder handmatige fijnregeling. Aan menskracht is in China echter geen gebrek, zo bleek mij iedere ochtend als drie Chinezen op één beeldscherm controleerden of het ontbijt wel bij mijn kamerprijs was inbegrepen.

weten sneller hoe ze die kunnen omzeilen dan dat hun ouders ze kunnen installeren.

Ik kreeg nu de kans de effectiviteit zelf uit te proberen vanuit mijn hotel in Peking (alle kamers voorzien van Internet, voor omgerekend vijf euro per dag). Eerst maar eens de Nederlandse Google. In een handomdraai was allerlei vnzigheid, politiek of anderszins, op het scherm te toveren. Waarschijnlijk vindt de Chinese overheid het gedachtegoed van Geert Wilders helemaal niet gevaarlijk voor

Column

Tom van Dijk

Brulapen-onderzoek



MARIE CÉCILE THIJS

'Sociaal-wetenschappelijk onderzoek met behulp van een computer gaat sneller, maar duurt langer'. Die stelling poneerde Rudy Andeweg bij het verschijnen van zijn proefschrift in 1982. Zijn these is lang waar geweest. Sinds een aantal jaar is zij evenwel genadeloos gefalsificeerd. Want online-onderzoek is een snelheidsduivel. Mooi natuurlijk: tijd is immers geld en geld is god.

Wat al héél snel gaat, moet alras natuurlijk nóg sneller. En wat steeds goedkoper is geworden, moet nóg goedkoper. Dat kan ook makkelijk, want rapporten worden niet in meer in het Nederlands geschreven, maar in Powerpoints – dat op korte termijn zal worden uitgeroepen tot de officiële onderzoekstaal. In dat idioom gelden twee kapitale stijlfiguren. Eén: hele zinnen zijn absoluut verboden. Twee: pictogrammen zijn verplicht. Zo brengt internet ons terug naar de Middeleeuwse plaatjesboeken.

Kort na het moment waarop het volk aan het fietsen kon raken, schreven dominees: 'de fiets doet kwaad'. Dat hadden ze goed gezien. Want de fiets bracht hun jongeren naar het Sodom en Gomorra van de urbaniteit. De fiets bracht het platteland evenwel ook naar de universiteit. Internet is ook een soort fiets. Iedereen kan er op stappen. Waar vroeger – voor veldwerk – nog de drempel lag van forse investeringen, is deze nagenoeg volledig weggefallen. Voor paar honderd euro is software te koop waarmee data zijn te produceren. En eenmaal geland in tabellen zien alle data er hetzelfde uit.

Een jaar of wat geleden liet een verzekeraar een 'duur' onderzoek naar een nieuw product doen. Dat product had iets te maken met de beurs. Zo'n 17% van de doelgroep bleek kooplustig. Het onderzoek was nog niet gehouden of de beurs kachelde in elkaar. Geen probleem: een slimme jongen bij onze verzekeraar heeft toen via software van 325 euro een aantal enquêtevragen op de website van het eigen bedrijf gezet. Spectaculair resultaat: een koopintentie van 50%. Na veel gesoebat is het dure onderzoek toch nog een keer herhaald. De koopintentie was geslonken tot 3%.

McKinsey&Company, Planet Internet, NRC Handelsblad, AD, MSN en FHV BBDO hebben onlangs voor de tweede keer in successie 21minuten.nl gelanceerd. In 2005 zeiden ze over zichzelf: 'het belangrijkste internetonderzoek naar de toekomst van Nederland'. En: 'wij zorgen er voor dat er naar u geluisterd wordt'. Dat klinkt nog eens bemoedigend: Fortuyn is weliswaar dood, maar dan hebben we altijd 21minuten nog. Onlangs is de tweede meting gepubliceerd in een keurig rapport, met hele zinnen, zonder Powerpoints. Via herwegingen is gecorrigeerd voor internetvertekening en 'zelfselectie onzuiverheid'. Met die laatste term hebben ze het dan over de vele miljoenen die niet op eigen initiatief 21 minuten wilden besteden aan het invullen van de vragenlijst. Dat is heel knap. Mensen die niet aan het onderzoek hebben meegedaan worden via een goocheltruc toch nog in het onderzoek gewogen. Vroeger kon dat nog niet. Wat er niet was, kon toen nog niet doorweging tevoorschijn getoverd worden. Geïnteresseerd? Die handige brulaap van onze verzekeraar heeft het recept.

Tom van Dijk is directeur Beleidsonderzoek bij Intomart GfK

Gelezen

European Strategy Forum for Research Infrastructures (ESFRI): European roadmap for research infrastructures – Report 2006; ISBN 92 79 02694 1
 ESFRI werd opgericht in 2002 om de Europese Commissie strategisch te adviseren over wetenschappelijke onderzoeksinfrastructuren. Aan deze eerste Europese 'roadmap' hebben zo'n duizend vooraanstaande onderzoeksexperts een bijdrage geleverd. De roadmap bevat 35 aanbevolen projecten. In de alfa- en gammawetenschappen zijn zes projecten benoemd: twee voor de alfawetenschappen, drie specifiek voor de gammawetenschappen en één dat beide gebieden bestrijkt. Online edition: <http://cordis.europa.eu/esfri/home.html>

Research Council United Kingdom, Economic and Social Data Service: Guidance on data management
 Although written for Rural Economy and Land Use (RELU) research projects, this publication contains valuable information on data management in general. This guide replaces the *Guide to good practice: data management 2005*. Online edition: www.data-archive.ac.uk/news/publications.asp

K. Breedveld, A. van den Broek, J. de Haan, L. Harms, F. Huysmans en E. van Ingen: De tijd als spiegel (Den Haag, Sociaal en Cultureel Planbureau, 2006); ISBN 90 377 02 83x
 Hoeveel uren werken we in Nederland? Wie zorgen voor het huishou-

den en de kinderen en hoeveel tijd besteden ze daaraan? Hoeveel vrije tijd hebben we, en wat doen we in die vrije uren? Kijken we televisie? Sporten we? En hoeveel tijd ruimen we in voor onze sociale contacten? In 2005 hielden 2200 Nederlanders gedurende een week in oktober hun activiteiten bij in een dagboekje. Sinds 1975 zijn om de vijf jaar vergelijkbare aantallen Nederlanders bereid geweest dat te doen. Dankzij hen is het mogelijk een beeld te schetsen van de tijdsbesteding van de Nederlandse bevolking over de periode 1975-2005. In De tijd als spiegel wordt de tijdsbesteding anno 2005 beschreven en vergeleken met die in 2000 en eerdere jaren. Online beschikbaar: www.scp.nl en zie ook: www.tijdbesteding.nl

Ineke A.L. Stoop: The hunt for the last respondent, nonresponse in sample surveys (Den Haag, Sociaal en Cultureel Planbureau, 2005); ISBN 90 377 02 155 (Ned. samenvatting)

Bij het analyseren van data of het publiceren van interessante uitkomsten wordt meestal weinig aandacht besteed aan de kwaliteit van de verzamelde gegevens. Geavanceerde analysemethoden kunnen nooit tekortkomingen in de dataverzameling goedmaken. Een hoge nonrespons kan grote gevolgen hebben voor de betrouwbaarheid van de resultaten. Daarom is investeren in de kwaliteit van de gegevens en kennis over de wijze waarop gegevens zijn verzameld van belang voor beleid en wetenschap. In Nederland is het steeds moeilijker om een hoge respons te krijgen. Dit



maakt onderzoek onder steekproeven uit de bevolking lastig en kostbaar. Een lage respons roept vragen op over de betrouwbaarheid van de uitkomsten. In dit proefschrift wordt een overzicht gegeven van de internationale literatuur over nonrespons in steekproefonderzoek. Online: www.scp.nl