

Printpedia?

Een beetje lacherig en ongeloofig was de algemene reactie toen het Duitse mediabedrijf Bertelsmann in 2008 een papieren versie van de Wikipedia aankondigde. Inmiddels is het boek verschenen. Al zijn er signalen dat het bij deze eerste uitgave blijft, het fenomeen is bijzonder genoeg om nader in ogenschouw te nemen.



Eerst wat feiten. Wat staat erin? Het boek bevat twintigduizend artikelen op 992 bladzijden; dat wil zeggen dat er ongeveer twintig artikelen op een pagina staan. De inhoud van de Duitstalige Wikipedia, die aan de tekst ten grondslag ligt, is dus in twee opzichten radicaal bekort: het aantal artikelen online (nu 940.000) is met ruim 97% teruggebracht; het aantal woorden per artikel is nog eens gereduceerd met meer dan 95 procent, zo leert een kleine steekproef. De koper koopt dus niet de encyclopedie, en zo wordt het boek dan ook niet verkocht. De titel luidt *Das Wikipedia Lexikon in einem Band. Ausgabe 2009*. Het bekorten van de artikelen is volgens een eenvoudige procedure gebeurd. Wikipedia-artikelen beginnen met een inleiding, vaak niet meer dan één of enkele zinnen. De uitgever heeft het

Wikipedia Lexikon samengesteld op basis van deze inleidingen. De geselecteerde ingangen zijn die welke online het vaakst worden opgevraagd. Het voorwoord van het *Lexikon* gebruikt dat als verkoopargument: het boek bevat niet alleen de begrippen uit een overgeleverde canon, maar ook films, popsterren en andere begrippen uit de cultuur van alledag.

Mag dat zomaar, het werk van vrijwilligers overnemen en daar winst mee maken? Het antwoord is duidelijk: ja, dat mag. Wikipedia-teksten kunnen door iedereen worden gebruikt, ook in aangepaste vorm, ook voor commerciële doeleinden, mits de originele auteur de credits krijgt voor zijn werk en het aangepaste product zonder beperkende copyrightbepalingen aan derden ter beschikking wordt gesteld. Deze voorwaarden leiden tot wat zonder twijfel de meest curieuze dertig pagina's uit het boek zijn: de pagina's waar in een zeer klein lettertype de in totaal ruim



99000 auteurs worden genoemd die aan de Wikipedia artikelen hebben bijgedragen. Wellicht in een poging om eventuele critici de wind uit de zeilen te nemen, kondigt de uitgever in de inleiding aan de Duitse Wikipedia financieel te zullen ondersteunen. Want sommige van de auteurs voelen zich wel degelijk gebruikt, zo

blijkt uit levendig webdebat.

Het *Lexikon* is dus legaal. Maar is het ook handig? Het lijkt onwaarschijnlijk dat iemand met toegang tot het Internet het boek zal raadplegen (want: niet doorzoekbaar, geen aanklikbare verwijzingen, artikel wellicht gedateerd). Misschien als de computer in een andere kamer staat, of er te lang over doet om op te starten? Het *Lexikon* zou dan concurreren met het netbook (kleine laptop, vooral voor huiskamergebruik). Een reden om het *Lexikon* aan te schaffen kan ook zijn dat onbewust een boek toch iets betrouwbaarder lijkt dan een website.

Heeft de redactie een betrouwbaarder boek willen maken dan de Wikipedia? Men geeft aan 'alleen zeer behoedzaam' de teksten voor het boek te hebben aangepast, maar geeft daarvoor tamelijk wijde criteria (bijvoorbeeld: belang van het onderwerp blijkt niet uit de tekst). Er wordt niet aangegeven in welke artikelen wijzigingen zijn aangebracht. Voor Wikipedia berust betrouwbaarheid juist mede op het fundamentele uitgangspunt dat precies kan worden nagegaan wie welke wijzigingen aanbrengt. De uitgever lijkt hier meer op zijn gezag te vertrouwen. Een uiting van dit autoriteitsdenken is ook dat men de aangebrachte veranderingen niet zelf in de Wikipedia-artikelen heeft willen doorvoeren maar ze 'ter controle aan de Wikipedia-gemeenschap ter beschikking gesteld'. Uit het woord 'controle' blijkt de achtergrond van de traditionele maker van naslagwerken.

Peter Boot

Column

Hans van Maanen

Enquetekeurmerk

Ook voor nieuws, en zeker voor nieuws over opiniepeilingen, geldt de wet van Gresham: slecht werk verdrijft goed werk. Dat ligt natuurlijk aan de journalisten, die goed werk niet van slecht werk kunnen onderscheiden, geen tijd hebben om de merites van een onderzoek te bepalen, en liever scoren dan horen en wederhoren. Of het ligt aan de opiniepeilers, die in hun streven naar media-aandacht niet altijd alle zwakke plekken en losse schroeven in hun onderzoek benadrukken.



BARBARA THIEL

Over deze kwestie kan men symposiaal lang discussiëren, en dat gebeurt dan ook regelmatig, maar mijn vraag is: kunnen bonafide onderzoekers wetenschapsjournalisten niet een beetje helpen het onderscheid tussen goed en slecht surveyonderzoek te maken? Zodat de onzin uit de krant blijft, en het goed onderbouwde nieuws een betere kans krijgt?

Wat is het belangrijkste verschil tussen wetenschap en opinieonderzoek? Als het over wetenschap gaat, weet elke journalist – en een beetje ervaren lezer – wel dat een verhaal in Nature een andere status heeft dan een verhaal in de Acta Zoologica Bulgarica. Het ene blad is het andere niet, en dat werkt als een soort keurmerk. Een dergelijk filter heeft het opinieonderzoek in Nederland niet. Er is al vaak over keurmerken gepraat, maar nog steeds mag rijp en groen tot advertentie/persbericht worden voorgekookt en aan de kranten worden opgediend.

Hier zie ik een mooie rol voor het nieuwe Nederlandstalig Platform voor Survey-Onderzoek. Als de leden van het platform nu eens voortaan in elk persbericht in de derde zin iets zetten als: 'Dit onderzoek is, naar de maatstaven van de branche-vereniging, redelijk betrouwbaar.' Of, als het beter is uitgevoerd, 'betrouwbaar'. Of, als er bijna niets op is af te dingen: 'zeer betrouwbaar'. In een voetnoot zou dan kunnen worden uitgelegd wat de criteria precies inhielden: was het een online-panelonderzoek of een echt deur-tot-deuronderzoek, is het betrouwbaarheidsinterval redelijk of smal, is het representatief, valide, dat soort dingen.

De meeste journalisten zetten zo'n persbericht toch zo in de krant, dus dan gaat het meteen goed, en van degenen die wat kritischer zijn, wordt eerst de nieuwsgierigheid geprikkeld en vervolgens nemen zij het waarschijnlijk als handzame informatie ook gewoon over. Het is een soort keurmerk, door de branche-organisatie zelf in het bericht gezet.

Het aardige is natuurlijk, dat onderzoek dat niet aan die minimumeis van 'redelijk betrouwbaar' voldoet – we weten allemaal over welke bureaus we het hebben – een dergelijke frase niet mee krijgt, en dus na verloop van tijd vanzelf door de mand valt. 'Hé', denkt de journalist, en de lezer, 'er staat niet bij of het betrouwbaar is, dan zal het wel niet deugen.'

Het lijkt mij, als wetenschapsjournalist, wel een mooi idee. De uitwerking ervan – wat de criteria dan worden, en wie dat gaat bepalen en controleren – laat ik uiteraard graag aan het platform over.

Hans van Maanen is freelance wetenschapsjournalist.

Gelezen

Madeleine de Cock Buning, Allard Ringnalda, Tina van der Linden (CIER): De juridische status van ruwe data; een wegwijzer voor de onderzoekspraktijk; Utrecht, SURFfoundation, juli 2009

Rapport om duidelijkheid te verschaffen over de juridische bescherming van onderzoeksdata, voor de onderzoeker die wil weten wat hij mag doen met de data van anderen. De mogelijkheden van ICT en internet maken de toegankelijkheid van onderzoekresultaten en -data steeds groter en opener. De mogelijkheden groeien om aanvullend aan tekst, publicaties te verrijken met o.a. de onderzoeksdata. Door deze toegankelijk te maken, wordt het makkelijker om de resultaten van onderzoek te verifiëren, te reproduceren en te hergebruiken. Maar met hergebruik van de (ruwe) onderzoeksdata is het belangrijk te weten wat de juridische status is van het materiaal. Soms is toestemming van de maker vereist maar sommige handelingen kunnen ook zonder toestemming worden verricht. SURFdirect wil inzicht verschaffen in de juridische bescherming van onderzoeksdata, voor de onderzoeker die wil weten wat hij mag doen met de data van anderen. Met deze informatie kan de onderzoeker ook bepa-

len hoe wenselijk bescherming van de eigen onderzoeksdata is. www.surfoundation.nl/nl/publicaties/Pages/Default.aspx



Inge Angevaere (red.): Toekomst voor ons digitaal geheugen, duurzame toegang tot informatie in Nederland; Den Haag, Nationale Coalitie Digitale Duurzaamheid (NCDD), juli 2009

Het is voor het eerst dat in Nederland de risico's die digitale bestanden lopen én de maatregelen die nodig zijn om die risico's te beheersen over de

hele breedte van de publieke sector (wetenschap, overheid, cultureel erfgoed) in kaart zijn gebracht. Het rapport besteedt vooral aandacht aan de organisatorische en beleidsmatige maatregelen die nodig zijn om te voorkomen dat ons digitaal geheugen onbereikbaar wordt omdat nieuwe computers de oude bestanden niet meer kunnen lezen. Het nu gepubliceerde document is een interimrapportage die besproken zal worden tijdens de nationale conferentie die de NCDD op 18 september 2009 organiseert. Na die consultatieronde zal de NCDD haar strategische plan vaststellen voor het realiseren van een duurzame infrastructuur voor toegang tot digitale informatie in Nederland. Het rapport werd geschreven door een projectteam bestaande uit René van Horik (Data Archiving and Networked Services, DANS: wetenschap), Annelies van Nispen (Digitaal Erfgoed Nederland, DEN: erfgoed), Ingrid Dillo (Koninklijke Bibliotheek, KB: projectmanagement), Inge Angevaere (NCDD: projectmanagement), Petra Helwig (Nationaal Archief, NA: overheid/archieven).

www.ncdd.nl/activiteiten-natverkenning.php

Paul Brous: Managing metadata for geographic information; Delft, TU Delft, GIMA Masters Thesis, 2009

Our knowledge of the real forces and mechanisms behind metadata production is still very limited. The reason for this difficulty might lie in the complex, dynamic and multifaceted interaction of organisations, the metadata producers themselves, the available technology, the spatial data and the metadata itself during the metadata production process. In order to be able to produce appropriate metadata we must have a proper understanding of what the main barriers and issues surrounding metadata production are and how these is-

sues affect the appropriateness of the metadata. This research describes a new approach to understanding the issues preventing appropriate metadata production by investigating them through the dimensions of a Geographic Information System (GIS) as described by Traub (1998) and comparing them to the categories of data quality as described by Strong et al. (1997).

www.msc-gima.nl/index.php/news