

INGEZONDEN

Wat is slecht onderzoek?

In het septembernummer van *e-data@research* pleitte Hans van Maanen voor een Enquêtekeurmerk. Hij stelde voor dat de leden van het Nederlandstalig Platform voor Survey-Onderzoek in ieder persbericht over een onderzoek vermelden hoe

van fervente internetgebruikers. Dat moet de onderzoeker dan wel duidelijk in zijn rapportage vermelden en de journalist in zijn artikel. Een onderzoek kan dus voor het ene doel slecht zijn en voor het andere doel misschien wel goed.

een grote random steekproef tot een foutief resultaat leiden. Lastig is dat verschillende componenten van kwaliteit elkaar niet kunnen compenseren. Als de vragen niet deugen, is de hoogte van de respons irrelevant. En als je de resultaten snel nodig hebt (Vond men dat er nieuwe verkiezingen zouden moeten komen als Balkenende president van Europa zou zijn geworden?) heb je geen tijd voor een uitgekiend responsverhogend programma.

Een keurmerk is dus niet zo makkelijk. De tweede suggestie van Van Maanen is eenvoudiger uitvoerbaar. Vermeld bij iedere dataverzameling de omvang van de steekproef, hoe de respondenten zijn geselecteerd, de wijze van data verzamelen en de nonrespons en geef een link zodat de geïnteresseerde lezer op haar gemak



NATIONALE BEELDBANK

betrouwbaar de gegevens zijn volgens de maatstaven van het NPSO, en in een voetnoot uitleggen hoe ze tot dit oordeel zijn gekomen.

Dit is niet zo makkelijk als het lijkt. De kwaliteit van een onderzoek heeft vele aspecten en hangt onder meer samen met het doel. Eigenlijk is het makkelijker om te zeggen wat slecht onderzoek is dan wat goed onderzoek is. Als je als respondent de volgende vraag voorgelegd krijgt, weet je bijvoorbeeld zeker dat het een slecht onderzoek is.



NATIONALE BEELDBANK

Vindt u dat de regering meer aandacht moet geven aan het snel afbouwen van de staatsschuld of meer geld besteden aan zaken voor het land en de burger?

Ja; Nee; Weet niet/geen mening

En als een onderzoek naar het gemak van elektronische belastingaangifte wordt uitgevoerd onder een steekproef van doorgewinterde internetgebruikers, is ook duidelijk dat dit geen representatief beeld geeft van de hele bevolking. Misschien geeft het wel een beeld van het oordeel

In de wereld van de surveys is langzamerhand overeenstemming bereikt over kwaliteitsmaatstaven. Een onderzoek moet het onderwerp goed dekken, de resultaten moeten accuraat zijn, de gegevens moeten tijdig beschikbaar komen en toegankelijk zijn en het gebruik van standaardvragen voor achtergrondgegevens wordt aanbevolen. De accuraatheid van resultaten neemt in het algemeen toe naarmate de steekproef groter is, maar dat geldt alleen bij een random steekproef. En helaas kan hoge nonrespons ook bij

de hele vragenlijst kan doorlezen.

Dat zou een belangrijke eerste stap zijn. In de tussentijd kunnen we binnen de NPSO verder discussiëren over nut en noodzaak van een keurmerk, en ook alvast een overzicht geven van alle bestaande richtlijnen en kwaliteitsoverzichten die we kennen. Wil je mee doen: meld je aan op www.npsso.net. We komen op dit onderwerp graag nog eens terug.

Ineke Stoop is mede-oprichter en lid van de kerngroep NPSO (www.npsso.net)

Column

Liesbeth Koenen

De hinderpaal taal

Droombeeld: alle wetenschappelijk onderzoek van alle tijden en talen staat on line. Liefst samen met lekker veel bronnen en ruwe gegevens.

Dus niet alleen boeken en artikelen, maar ook kleitabletten en papyrusrollen, oude dagboeken en lab-aantekeningen, en nog oneindig veel meer. En dan tik of spreek je wat zoektermen in, en floep, daar verschijnt in je eigen taal wat de Japanners erover te zeggen hebben, wat er in Engeland in de fameuze zeventiende eeuw over is opgeschreven en welke conclusies Indiase onderzoekers trekken. Je komt vanzelf uitsluitend terecht bij gegevens die ertoe doen, en overal vind je handzame samenvattingen.

Nu de echte wereld. Een Telegraafbericht op internet over een zware zeebeving toont direct daarnaast een advertentie voor een fijn cursusweekend overleven op de Rucphense heide. Laatst vond een vriendin een brief op de mat, gericht aan 'Lieve inwoners van de busje Eeghenlaan en het busje Eeghenstraat'. Ook de rest van het schrijven ('de inbreker ging door onze rug tuin', 'verhuur alstublieft ons weten') begon pas enigszins begrijpelijk te worden toen ze zich realiseerde dat het om een computervertaling uit het Engels ging (van is busje, back achter en rug, en let zowel laat als verhuur).

De automatische omzetting van achttiende-eeuwse artikelen uit De Vaderlandsche Letteroefeningen met Optical Character Recognition (te vinden bij e-laborate.nl) maakt van 'het menschelyk geslagt' 'het menfchelyk geflagt' en achter 'geraoedsgefeldheid' blijkt bij controle gewoon 'gemoedsgesteldheid' schuil te gaan. Verbaasd lees ik deze week in een mailtje 'Ok zit Nietzsche boot niks op 40000 vortex te mailen.' De volgende dag volgt gelukkig de uitleg: het is wat je na autocorrectie van iPhone overhoudt van 'Ik zit niet voor niks op 40000 voet te mailen.' Spellingcheckers vinden tegenwoordig flessentelefoon, geelhork en opmuizen prima woorden. En ook tegen een computer praten, levert nog altijd hilarische resultaten op.

Ze liggen voor het opscheppen. Bij de weidse vergezichten op een ideale digitale wereld staat kortom een ding gigantisch in de weg: taal. Computers begrijpen er nog steeds geen bal van. Intussen wordt ons al sinds 1945 'binnen vijf jaar' de vertaalcomputer beloofd. Er wordt gesleuteld en gedaan, en soms lijkt het nog heel wat. Maar het fundament ontbreekt. We weten zelf veel te weinig over ons ingenieuze taalvermogen.

Het rare is intussen dat alleen taalkundigen dat lijken te beseffen. Taal is ons kennelijk zo eigen, is zo gewoon, dat het voor niet-ingevoerden meestal een makkie lijkt. Of weet iemand een betere verklaring voor het onbegrijpelijk feit dat taalonderzoek niet allang op elke universiteit vooraan staat zodra de onderzoeksgelden worden verdeeld?

Liesbeth Koenen is taalkundige en wetenschapsjournalist



MARCO FRIGERIO

Gelezen

Peter Boot: Mesotext. Digitised Emblems, Modelled Annotations and Humanities Scholarship; Amsterdam, Amsterdam University Press, 2009; Proefschrift Universiteit Utrecht, november 2009; ISBN 978-90-89641-87-8

De belangrijkste ontbrekende functionaliteit in huidige digitale edities is die van annotatie. Digitale edities zouden aan onderzoekers de mogelijkheid moeten bieden om gestructureerde en ongestructureerde observaties met betrekking tot de uitgegeven tekst vast te leggen. Dit boek bespreekt een aantal benaderingen van annotatiesystemen. Het doet dat in het kader van de studie van het embleem, het zestiende en zeventiende eeuwse literair genre waarin een afbeelding, een motto en een vaak moraliserende epigram met elkaar verbonden werden. Bij een juiste aanpak kan annotatie uitgroeien tot mesotekst: tekst gepositioneerd tussen de geannoteerde teksten en de wetenschappelijke artikelen en monografieën waarvoor de annotaties de argu-

menten leveren. In een digitale context moet het mogelijk zijn om heen en weer te navigeren tussen geannoteerde tekst, annotatie en artikel.

M. Senten (redactie): Experiment NL 2, wetenschap in Nederland; Diemen/Den Haag, G+J Publishing, 2009; ISBN 978-94-6044-014-4

Het vierde NWO-publieksboek Experiment NL 2 beschrijft in begrijpelijke taal de meest opmerkelijke, briljante, leuke en spannende onderzoeksprojecten van het afgelopen jaar. Onderzoekers uit alle disciplines komen aan het woord en tonen hun bevindingen aan mensen die geïnteresseerd zijn in wetenschap, vernieuwing en de wereld om hen heen. *Verwondering, Op onderzoek en Experiment NL deel 1* gingen deze uitgave voor.

Henk Vinken: Verkenning Jeugddata. Raadplegingen van jeugdonderzoekers met betrekking tot hun datawensen. Den Haag, Data Archiving and Networked

Services (DANS), 2009; Studies in Digital Archiving 3; ISBN 978-94-90531-01-0

Jeugd heeft nu veel aandacht in wetenschap, overheidsbeleid en samenleving. Er wordt veel jeugdonderzoek gedaan door diverse onderzoeksinstituten en universiteiten, zowel landelijk als regionaal. Tot op heden is er echter weinig systematisch inzicht in aard en kwaliteit van dit jeugdonderzoek en van de onderliggende onderzoeksgegevens. DANS (Data Archiving and Networked Services) heeft daarom in de eerste helft van 2009 een Dataverkenning Jeugd uitgevoerd. Deze publicatie, uitgebracht in de reeks Studies in Digital Archiving, doet verslag van die verkenning.

Karl Fogel: Open source-software produceren, met succes een open source-softwareproject runnen; Utrecht, Stichting Kennisnet, SURFfoundation, SURFdiensten en SURFnet; Juli 2009

Het boek *Producing Open Source Software* van de Amerikaanse Open Source-des-



kundige Karl Fogel geldt in de internationale Open Source-community als een standaardwerk en is inmiddels in diverse talen beschikbaar. Sinds afgelopen zomer ook in het Nederlands. SURF en Kennisnet hebben de vertaling, met de titel *Open Source-software produceren*, financieel mogelijk gemaakt.

Het boek is bedoeld voor ontwikkelaars en managers van open source die overwegen een open source-project te beginnen, en voor degenen die al een project gestart zijn en zich afvragen hoe ze verder moeten. Het kan ook handig zijn voor mensen die alleen willen participeren in een open source-project, maar dit nooit eerder gedaan hebben. De lezer hoeft geen programmeur te zijn, maar moet wel de technische basisconcepten van software kennen, zoals 'broncode', 'compilers' en 'patches'. Ervaring met open source-software, als gebruiker of ontwikkelaar, is niet nodig. Mensen die al aan open source-softwareprojecten hebben meegewerkt, zullen sommige delen van het boek waarschijnlijk als een open deur beschouwen en deze waarschijnlijk willen overslaan. Omdat er grote verschillen kunnen zijn wat betreft de ervaring die de lezers hebben, zijn de secties van duidelijke koppen voorzien en wordt vermeld wanneer iets overgeslagen kan worden door degenen die al bekend zijn met deze materie. De pdf-versie van de vertaling is kosteloos beschikbaar: www.surf.nl/nl/OverSURF/Publicaties