

Vluchtig medium vraagt om degelijke analyse: methodologische overwegingen bij de analyse van internetmateriaal

Martine van Selm

Samenvatting

In deze bijdrage wordt stilgestaan bij een aantal methodologische uitdagingen bij het analyseren van webmateriaal. Dit gebeurt aan de hand van achtereenvolgens:

- a. typen vraagstellingen in web onderzoek;
- b. het selecteren van onderzoeksmateriaal;
- c. het definiëren van eenheden; en
- d. het coderen en archiveren van webmateriaal.

Inleiding

Sociaalwetenschappelijk onderzoek naar en met behulp van het internet wordt steeds gangbaarder. Er zijn een aantal kenmerken van het World Wide Web (WWW), die relevant zijn voor de vraag hoe webmateriaal onderzocht kan worden (Weare & Lin, 2000): interactiviteit, het netwerkarakter van webmateriaal, de multimediale elementen van boodschappen, en het wereldwijd bereik. Een innovatieve benadering van webmateriaal gaat uit van het dynamisch karakter en de zogenoemde *hypertextualiteit van webmateriaal* (Schneider & Foot, 2005). Websites zijn in deze opvatting geen afgebakende eenheden zoals krantenberichten of personen, maar zijn door hyperlinks eindeloos met elkaar verbonden. Dit netwerk geeft wat Schneider en Foot noemen de *websfeer* een eigen karakter. In deze bijdrage worden vier methodologische uitdagingen bij het analyseren van webmateriaal besproken: a) het type vraagstellingen in onderzoek naar webomgevingen, b) het selecteren van onderzoeksmateriaal, c) het definiëren van eenheden, en d) het coderen en archiveren van internetmateriaal.

Vraagstellingen

Onderzoek naar webmateriaal is vaak gedaan met behulp van vraagstellingen van exploratief-beschrijvende aard. Een voorbeeld is de vraag of digitale kranten en websites van televisienieuwsstations gebruikmaken van de interactieve mogelijkheden van het internet. Of de vraag hoe de actieve of passieve rol van deelnemers aan een politieke discussie op internet gestalte krijgt (Hagemann, 2002). Maar ook studies vanuit meer theoretisch georiënteerde vraagstellingen weten internetmateriaal te gebruiken. Een voorbeeld hiervan is het onderzoek van Bordia en Rosnow (1998) naar de verspreiding van geruchten binnen een online discussieomgeving waarin gebruik wordt gemaakt van het netwerkarakter van het internet om de verspreiding van een gerucht *unobtrusively* te kunnen

volgen. Andere studies gaan na of bekende kenmerken van bijvoorbeeld (politieke) discussies, ook op internetmateriaal van toepassing zijn (bijvoorbeeld Van Selm, Jankowski & Tsaliki, 2002); of zijn gericht op de manier waarop bestaande theorievorming rondom zingeving toegepast kan worden op zingevingpraktijken op persoonlijke homepages (Van Selm & Hijmans, 2006).

Omdat via computernetwerken (groepen) mensen onderling met elkaar verbonden worden, zijn deze te beschrijven als sociale netwerken (Felling & Van Selm, 2006) en te onderzoeken met behulp van sociaal netwerkanalyse. Binnen de netwerkanalyse fungeert de relatie tussen actoren als de analyse-eenheid en worden aan deze eenheid kenmerken onderscheiden zoals aard, gerichtheid en sterkte. Voorbeelden van de *aard van een relatie* binnen een internetcontext zijn het delen van informatie, het bediscussiëren van werk, het geven van emotionele steun of het bieden van gezelschap. Bij *gerichtheid van een relatie* gaat het over de vraag wie de donerende dan wel ontvangende partij is, terwijl bij *sterkte van een relatie* gedacht kan worden aan zaken als frequentie van contact, het belang of de hoeveelheid van datgene dat wordt uitgewisseld. Binnen onderzoek naar computer gemedieerde communicatie (CMC) netwerken kunnen ten minste twee type onderzoeksvragen worden onderscheiden: vragen naar kenmerken van CMC-netwerken en vragen naar de rol van CMC ten opzichte van andere communicatiekanalen voor het opbouwen en in stand houden van een sociaal netwerk. Ten behoeve van het eerste type vraag kunnen logdata gebruikt worden. Logdata zijn gegevens die door communicatienetwerken zelf automatisch gegenereerd worden en die een directe inzicht geven in communicatiepatronen.

Dit type data werd gebruikt in een onderzoek naar het gebruik van een aan een academisch ziekenhuis verbonden digitale poli. Deze digitale poli is een website bedoeld voor specifieke groepen patiënten die binnen een Nederlands academisch ziekenhuis een behandeling ondergingen. Op de website is zowel persoonlijke patiëntinformatie te vinden, als algemene informatie over aandoening en behandeling. Daarnaast biedt de digitale poli mogelijkheden tot consultatie van professionals en medepatiënten. Rondom de poli werd een studie verricht naar het gebruik van de verschillende websiteomgevingen door patiënten gedurende de opeenvolgende fases van de behandeling (Tuil, Van Selm, Verhaak, De Vries & Kremer, 2009) en een studie naar de onderwerpen waarover gepraat wordt door patiënten en professionals binnen de chatruimte van de digitale poli (Van Selm, Tuil, Verhaak, Woldringh & Kremer, 2008).

Materiaalselecties samenstellen

Steekproeven kunnen dankzij de toegankelijkheid van materiaal via het internet veel groter zijn dan voorheen. Ook is het vanwege zoek- en selectiemogelijkheden die het internet aanbiedt makkelijker dan voorheen om specifieke steekproeven samen te stellen. Een voorbeeld hiervan is de databank LEXIS-NEXIS (www.lexisnexis.nl) waarin, onder andere, op onderwerp gezocht kan worden naar artikelen uit een groot aantal nationale en internationale dagbladen. Andere voordelen van het internet zijn dat materiaal ontsloten wordt dat vroeger te duur was om te verzamelen of dat vroeger buiten het gezichtsveld

van onderzoekers viel. Daarnaast is, zoals gezegd, het internet zelf ook een platform voor veel nieuw materiaal waar inhoudsanalyses op uitgevoerd kunnen worden.

De omvang en veranderlijkheid van het internet bemoeilijkt het trekken van een aselechte steekproef die aan de kwaliteitscriteria van sociaalwetenschappelijk onderzoek voldoet. McMillan (2000) wijst hierbij op het feit dat het aantal websites continu groeit en dat indexen incompleet zijn en elkaar overlappen. Weare & Lin (2000) beschrijven een aantal hulpmiddelen die te gebruiken zijn bij het selecteren van een steekproef van webmateriaal. Deze hulpmiddelen betreffen het gebruik van a) domeinnamen als steekproefkader, b) zoekmachines en c) (persoonlijke) verzamelingen van websites en populaire sites.

Onderzoekers die hun steekproefkader willen opbouwen op basis van internetadressen zoals domeinnamen, moeten rekening houden met de hiërarchische structuur van het Domein Name System (DNS) en de decentrale manier waarop dit systeem wordt beheerd. Weare en Lin (2000) onderscheiden drie domeinniveaus. Ten eerste bestaat het *top-level domein*. Voorbeelden uit een Amerikaanse context zijn .org, .edu, .int. Andere voorbeelden zijn afkortingen van namen van landen zoals .nl, of .de. Daarnaast kan een domein van een tweede orde niveau (bijvoorbeeld usc.edu) worden onderscheiden. Aan de hand van een tweede orde niveau kunnen specifieke groepen van Britse commerciële sites (uk.com) worden onderscheiden, of bijvoorbeeld zoekmachines (yahoo.com). Aan de hand van domeinen van de derde orde kunnen internetadressen verder verfijnd worden (bijvoorbeeld annenberg.usc.edu).

Het DNS-systeem kent een aantal beperkingen. Het systeem heeft geen centraal beheer waardoor het gebruik van de hiërarchische structuur zeer varieert. Daarnaast zal een steekproef die getrokken is op basis van DNS geen websites bevatten die genesteld zijn binnen de directories van iemands computer. Toch vinden Weare en Lin (2000) dat domeinnamen een bruikbaar houvast bieden bij het trekken van een steekproef, juist vanwege de hiërarchische structuur en de relatieve compleetheid van het DNS-systeem. Dit houvast geldt met name als het gaat om globale onderzoeksvragen over de inhoud van webmateriaal, en niet om onderzoeksvragen die gaan over aspecten van websites die buiten het bereik van het DNS-systeem vallen, zoals van afdelingen binnen grotere organisaties. Het DNS-systeem is, kortom, bruikbaar als het gaat om onderzoeksvragen die relevant zijn voor categorieën internetmateriaal die te vinden zijn via internetadressen die op domeinen gebaseerd zijn.

Een tweede hulpmiddel bij het opbouwen van een steekproefkader wordt gevormd door *zoekmachines*, een gemakkelijk en snel te gebruiken hulpmiddel. De selectie via de zoekmachines vindt plaats op het niveau van het topdomein. Het belangrijkste nadeel is de onsystematische en selectieve manier waarop procedures voor de indexering van websites worden toegepast door de bedrijven achter de zoekmachines (Hargittai, 2007). Dit laatste kan resulteren in steekproeven die sterk vertekend zijn in het voordeel van de websites die veel internetverkeer genereren. Deze vertekening is op zichzelf niet problematisch in studies naar de inhoud van sites die een gemiddelde websurfer zal tegenkomen. Maar in studies die gericht zijn op het in kaart brengen van de inhoud van een ander type

sites, zoals sites van besloten of exclusieve gemeenschappen, is dit een serieuzer probleem. Een remedie tegen deze vertekening is het gebruik van zogeheten metazoekmachines, zoals MetaCrawler (www.metacrawler.com). Deze zoekmachines genereren lijsten van websites die gebaseerd zijn op de zoekacties van meerdere zoekmachines tegelijkertijd. Een voordeel hiervan is dat de zoekacties completer worden; een nadeel blijft dat ze nog steeds de meest bezochte sites vinden en in die zin vertekening niet tegen gaan.

Als derde hulpmiddel bij het vormen van een steekproefkader kunnen *verzamelingen van sites* genoemd worden, zoals populaire sites (in de betekenis van veelbezochte) of (persoonlijke) portal pages. In verzamelingen worden websites rondom een bepaald onderwerp veelal geordend gepresenteerd. Het voordeel van het gebruik van verzamelingen als steekproefkader is dat deze sites vaak meer 'on topic' zijn dan sites die via zoekacties met zoekmachines op het spoor zijn gekomen. Het nadeel is echter dat verzamelingen altijd incompleet zijn, in grote mate willekeurig en daarmee vertekend in vaak onbekende richting (Weare & Lin, 2000). Als men kiest om een lijst van populaire sites te gebruiken als steekproefkader wordt vaak de hulp ingeroepen van een internettool die het verkeer naar populaire sites bijhoudt (bijvoorbeeld www.10hot.com) of tools die in kaart brengen welke zoektermen vaak gebruikt worden door internetgebruikers (bijvoorbeeld www.searchword.com).

Definiëren van onderzoekseenheden

In inhoudsanalyses van webmateriaal wordt soms als waarnemingseenheid de gehele website gekozen. Weare en Lin (2000) waarschuwen hier echter tegen omdat het onrealistisch is om van codeurs te verwachten dat zij een website als geheel betrouwbaar kunnen coderen. Van Selm, Jankowsli en Tsaliki (2002) definieerden als waarnemingseenheid de webpagina's die aanwezig zijn op de eerste drie 'lagen' van de websites van drie Nederlandse politieke partijen (CDA, GroenLinks, en de Socialistische Partij). Dit onderzoek was gericht op het in kaart brengen van hoe drie centrale elementen van het democratisch proces (ontsluiten van informatie, deliberatie en inspraak bij besluitvorming) terug te vinden zijn in de inhoud van de genoemde websites. Het onderzoeksmateriaal bestond, zoals gezegd, uit de eerste drie 'lagen' van webpagina's die toehoren tot de websites met URL-stam www.groenlinks.nl, www.sp.nl, en www.cda.nl. Op de eerste laag bevindt zich de welkomst- of startpagina van de website. Op de tweede laag bevinden zich alle pagina's die met één 'muisklik' vanuit de welkomstpagina te bereiken zijn; het zijn daarmee die webpagina's die bezoekers tegenkomen als ze vanaf de welkomstpagina klikken op een hyperlink of button anderszins. De derde laag wordt gevormd door alle pagina's die via een link geopend kunnen worden vanaf een van de pagina's op de tweede laag.

Pagina's op de tweede en derde laag hebben de URL-stam gemeen met de URL waarmee de beginpagina gelokaliseerd is op het WWW. Pagina's die wel bereikbaar waren via de eerste of tweede laag, maar buiten het domein van de URL-stam vielen (externe links), werden buiten beschouwing gelaten. Het onderzoek werd beperkt tot de

genoemde drie lagen van de website omdat het aantal webpagina's exponentieel groeit met elke 'laag' die eraan wordt toegevoegd. De beslissing om de websites op deze hiërarchische manier te analyseren kan bekritiseerd worden. De inhoud van een website kan zich immers ook op andere wijze aan de bezoeker presenteren. Websitebezoekers zijn in staat om pagina's die tot een bepaalde site behoren te bereiken via diverse 'zij-ingangen', bijvoorbeeld via 'deep links' vanuit andere websites. Toch werd verondersteld dat het leeuwendeel van de bezoekers 'aan de voorkant' de website binnenkomt. Deze veronderstelling werd ondersteund door informatie over logdata: voor de CDA website geldt dat ruim 75 procent van alle bezoekers door de website navigeert, beginnend met de eerste pagina (Van Selm, Jankowski & Tsaliki, 2002).

Weare en Lin (2000) beschrijven een aantal manieren om de grootte van de registratie-eenheid te reduceren, zoals het definiëren van frames binnen een webpagina of het definiëren van typen van webpagina's. Met betrekking tot het samenstellen van een selectie te analyseren materiaal wijzen deze auteurs op de waarde van gebruiksgegevens. Op basis van de hoeveelheid internetverkeer die een pagina naar zich toetrekt (via bijvoorbeeld het aantal pageviews op de teller van een webpagina) kan een onderzoeker besluiten pagina's die nooit of nauwelijks bezocht worden buiten beschouwing te laten in de inhoudsanalyse. Hetzelfde principe kan dienen als hulpmiddel bij het definiëren van context eenheden:

'If many requests for the webpages under analysis are referred from weblocations outside the context units, the scope of the context unit may need to be expanded.'
(Weare & Lin, 2000: 284)

De auteurs benadrukken het belang van software voor het monitoren van internetgebruik:

'Improvements in web tracking and analysis software, especially software that is user based rather than site based is needed to clarify the boundaries of messages in a hyper-text environment, thereby clarifying the proper definition of sampling and context units of analysis.'
(Weare & Lin, 2000: 289)

Hagemanns onderzoek naar bijdragen aan een politieke listserv discussielijst geldt als een goed voorbeeld van het expliciet benoemen van de eenheden:

'In the content analysis, the recording unit was the individual posting. Quotations from former contributions were excluded from the analysis, but could provide the context for the analysed part. The unit of analysis was the discussion list as a whole.'
(Hagemann, 2002: 62)

In het eerdergenoemde onderzoek naar de inhoud van chatberichten (Van Selm et al., 2008) werd het individuele chatbericht gehanteerd als eerste waarnemingseenheid en de

gehele chatsessie als tweede waarnemingseenheid. Ten behoeve van dit onderzoek werden alle chatsessies die binnen een bepaalde periode plaats hadden gehad (in totaal 15.304 chatberichten). De onderzoekers voerden een kwalitatieve inhoudsanalyse uit op een selectie van chatberichten, door middel van het open en axiaal coderen van chatberichten en chatsessies. In de chatsessies samen participeerden 33 patiënten en tien gezondheidszorgprofessionals. Een gemiddelde chatsessie duurde een tot anderhalf uur. In driekwart van de chatsessies deed een professional mee; zij domineerde de sessies echter niet: minder dan 25 procent van alle chatberichten was van hen afkomstig. In een gemiddelde chatsessie worden enkele regels gependend aan het begroeten van elkaar waarna vrij snel wordt overgegaan tot het uitwisselen van nieuwtjes over persoonlijke ervaringen die met de behandeling te maken hadden. Relatief veel tijd wordt ingeruimd voor het afscheid nemen van elkaar. Bij het samenstellen van een selectie uit dit chatmateriaal werd de homogeniteit van het materiaal in ogenschouw genomen. Als een populatie of universum aan materiaal geheel homogeen is, zou de analyse van een enkele case (hier een chatsessie) voldoende zijn om een beeld te schetsen van het te onderzoeken thema (hier de onderwerpen besproken in de chatsessie). Omdat van homogeniteit meestal geen sprake is bij sociale fenomenen, worden factoren benoemd die variatie zullen veroorzaken. De inhoud van de chatsessies varieert naar verwachting op basis van groepssamenstelling en behandelfase van de patiënten. Daarom werden chatberichten geproduceerd door vier verschillende combinaties van patiënten, en verschillende participerende professionals, verspreid over de fasen van de behandeling geselecteerd. Werkend vanuit het principe van verzadiging, dat wil zeggen dat het verzadigingspunt bereikt is als de analyse van nieuwe data geen nieuwe inzichten meer oplevert, werden per groep vijf sessies geanalyseerd. Deze twintig sessies representeerden 4042 chatberichten. Ten behoeve van het beantwoorden van de eerste onderzoeksvraag (*Over welke onderwerpen wordt gechat?*) fungeerde het individuele chatbericht als waarnemingseenheid. Ten behoeve van het beantwoorden van onderzoeksvragen 2 (*Op welke wijze spreken deelnemers over de behandeling?*) en 3 (*Op welke wijze spreken deelnemers over het omgaan met ongemakken behorend bij de behandeling?*) werd de gehele chatsessie gebruikt als waarnemingseenheid. Dit resulteerde in drie clusters van open codes verwijzend naar elk afzonderlijk chatbericht dan wel de gehele chatsessie. Door deze codes binnen de clusters en tussen de clusters onderling met elkaar te vergelijken werden categorieën geformuleerd. De categorieën vormen bouwstenen voor de antwoorden op elk van de onderzoeksvragen. De analyse werd gedaan met behulp CAQDAS-pakket Kwalitan (www.kwalitan.net/engels/index.html).

Archiveren en coderen

Hoewel in elke inhoudsanalyse het te onderzoeken materiaal gereedgemaakt moet worden alvorens met het daadwerkelijke coderen begonnen kan worden, verdient het veiligstellen (archiveren, ordenen, downloaden) van webmateriaal speciale aandacht. Een belangrijk gevolg van de hypertextuele aard van webmateriaal is dat websites decentraal beheerd worden en daarmee gaan en komen. Bij het coderen van webmateriaal moet hiermee rekening gehouden worden. Van belang is om datum en tijdstip waarop de

websites gecodeerd werden zorgvuldig te registreren, of nog beter: websites van een bepaalde datum te archiveren (Weare & Lin, 2000). Een andere praktische hint is dat onderzoekers zich ervan vergewissen dat alle codeurs precies hetzelfde materiaal voor zich hebben, dus dat zij hetzelfde type browser gebruiken, alsmede hetzelfde type computerscherm en internetverbinding. Controle hierop is met name relevant als codeurs vanwege de mogelijkheden die het internet hen biedt, op afstand van elkaar codeerwerkzaamheden verrichten. Ook is het belangrijk dat duidelijke instructies gegeven worden ten aanzien van de 'route' die gevolgd moet worden bij het navigeren van het te coderen webmateriaal. Dit alles ten behoeve van het verkrijgen van acceptabele waarden van de intercodeursbetrouwbaarheidscoëfficiënt. Sommige auteurs (o.a. McMillan, 1999) benadrukken juist in verband met het veranderlijke karakter van websites het belang van het monitoren over tijd van inhouden van websites.

In de eerdergenoemde studie naar de inhoud van chatberichten werden alle chatsessies die binnen een bepaalde periode plaatshadden gearchiveerd. Veranderlijkheid van het materiaal vormde daar dus geen bedreiging van de kwaliteit van het onderzoek. Gearchiveerd webmateriaal is in het algemeen minder gevoelig voor verandering omdat dit materiaal bewaard wordt in een aparte database die vaak alleen door onderzoekers met toegang geconsulteerd kan worden. Hetzelfde geldt voor de inhoudsanalyse van zogeheten 'stand alone' digitale documenten zoals computergames (Dietz, 1989). Er bestaan grootschalige archieven van websites die toegankelijk zijn voor onderzoekers, zoals een archief met sites gerelateerd aan Nederlandse politieke partijen (www.web-archivist.nl), of een archief met sites die allemaal gaan over de gebeurtenissen in de VS op 11 september 2001 (<http://september11.archive.org/>). Een ander grootschalig archief voor webmateriaal is www.archive.org. Het opbouwen van dergelijke grootschalige archieven gebeurt met behulp van speciale software en vraagt een actief beheer. Dit valt meestal buiten de reikwijdte van een individueel onderzoeksproject. Er bestaan echter wel verschillende kant-en-klare softwarepakketten waarmee websites gedownload en bewaard kunnen worden in een geheugen. TeleportPro (www.tenmax.com/teleport/pro/home.htm) is een voorbeeld van een dergelijk product dat waardevol kan zijn voor basale vormen van inhoudsanalyse. Andere voorbeelden zijn: <http://nodexl.codeplex.com>, www.tweetscan.com, <http://linkanalysis.wlv.ac.uk/>, <http://socscibot.wlv.ac.uk/>, www.mediaconverter.org, www.httrack.com, www.zotero.com.

In een onderzoek naar de inhoud van persoonlijke homepages (Van Selm & Hijmans, 2006) werden de homepages niet gearchiveerd. In plaats daarvan werden ze gelinkt aan een website van een groupware product (BSCW, <http://bscw.gmd.de/>) waartoe alleen de onderzoekers toegang hadden. Tijdens het onderzoek konden de te analyseren websites gemakkelijk opgevraagd worden. Omdat alleen de URL's van de homepages op de site van het groupware product geplaatst werden, lijkt het opvragen van de sites sterk op het opvragen van sites via een gewone webbrowser. Een extra handigheid van het groupware product voor met name verspreid gestationeerde onderzoeksteams is dat commentaar en tekstfragmenten toegevoegd kunnen worden aan de URL's. Deze manier van 'veilig stellen' van materiaal is bekritiseerbaar en in feite alleen zinnig als de onderzoeksperiode niet veel tijd beslaat en er niet teveel veranderingen in de samenstelling van het mate-

riaal zijn te verwachten. Immers, zodra een website verdwijnt van het WWW zal de link vanuit het groupware product naar de site niet meer functioneren.

Slot

In deze bijdrage werd stilgestaan bij methodologische uitdagingen bij het analyseren van webmateriaal. Dit gebeurde aan de hand van een bespreking van achtereenvolgens soorten vraagstellingen in onderzoek naar de inhoud van webomgevingen, het samenstellen van selecties aan onderzoeksmateriaal, het definiëren van eenheden en coderen en archiveren van webmateriaal. Bij de bespreking werd duidelijk dat gangbare methodologische principes van sociaalwetenschappelijk onderzoek richtlijnen bieden voor onderzoek naar webmateriaal, maar dat de aard van het webmateriaal, denk aan hypertextualiteit en vluchtigheid, onderzoekers voor nieuwe uitdagingen stelt.

Literatuur

Bordia, P. & Rosnow, R. (1998). Rumor rest stops on the Information Highway. Transmission patterns is a Computer-mediated Rumor Chain. *Human Communication Research*, 25(2), 163-179.

Dietz, T. (1998). An examination of violence and gender role portrayals in video games: Implications for gender socialization and aggressive behavior. *Sex Roles*, 38(5/6), 425-442.

Felling, A.J.A. & Selm, M. van. (2006). Netwerkanalyse. In F. Wester, K. Renckstorf & P. Scheepers (Red.), *Onderzoekstypen in de Communicatiewetenschap* (pp. 181-204). Alphen aan den Rijn: Kluwer.

Hagemann, C. (2002). Participation in and content of two Dutch political party discussion lists on the Internet. *Javnost/The Public*, 9(2), 61-76.

Hargittai, E. (2007). The social, political, economic, and cultural dimensions of search engines: An introduction. *Journal of Computer-Mediated Communication*, 12(3), article 1. <http://jcmc.indiana.edu/vol12/issue3/hargittai.html>.

McMillan, S.J. (2000). The microscope and the moving target: the challenges of applying content analysis to the World Wide Web. *Journalism and Mass Communication Quarterly*, 77(1), 80-98.

Schneider, S.M. & Foot, K. (2005). Web Sphere Analysis: An approach to studying online action. In C. Hine (Ed.), *Virtual methods: Issues in Social Research on the Internet*. Oxford: Berg Publishers.

Selm, M. van & Hijmans, E. (2006). Digitale documenten. Zingeving online: een exploratie van persoonlijke webpages. In F. Wester (Red.), *Inhoudsanalyse: theorie en praktijk* (pp. 207-226). Alphen aan den Rijn: Kluwer.

Selm, M. van & Jankowski, N.W. (2006). Conducting online surveys. *Quality & Quantity*, 40(3), 435-456.

Selm, M. van, Jankowski, N.W., & Tsaliki, L. (2002). Political parties online: Digital democracy as reflected in three Dutch political party Web sites. *Communications: The European Journal of Communication Research*, 27(2), 189-210.

- Selm, M. van, Tuil, W., Verhaak, C., Woldringh, G. & Kremer, J.A.M. (2008). Chat about what matters most: An analysis of chat contributions posted to an outpatient fertility website. *CyberPsychology and Behavior*, 11(6), 675-677.
- Tuil, W., Selm, M. van, Verhaak, C., Vries, P.F. de & Kremer, J.A.M. (2009). Dynamics of internet use during the stages of IVF. *Fertility and Sterility*, 91(3), 953-956.
- Weare, C. & Lin, W. (2000). Content analysis of the World Wide Web. Opportunities and Challenges. *Social Science Computer Review*, (18)3, 272-292.