Privacy en zoekmachines: vergezocht?

Mr. R.C. Winkelhorst

Roderic C. Winkelhorst is werkzaam bij het College bescherming persoonsgegevens.

Trefwoorden: *zoekmachine, bescherming van de persoonlijke levenssfeer, privacy, aansprakelijkheid, Google*

Via Google, de meest gebruikte zoekmachine, kan men interessante dingen vinden over allerlei bekende en geheel onbekende mensen die men zonder zoekmachine wellicht nooit zou hebben gevonden. Steeds meer mensen worden via die weg dan ook geconfronteerd met (ongewilde) publicaties van hun persoonsgegevens op het internet. Verhaal halen bij degene die de informatie op internet heeft gezet blijkt in de praktijk een moeilijke zaak. Vraag is echter of dit nu altijd wel noodzakelijk is. Door de razendsnelle ontwikkeling van diverse zoekdiensten en de representatie van informatie lijkt er een nieuwe privacydimensie op internet te zijn ontstaan. Die nieuwe privacydimensie heeft mogelijk gevolgen voor de rol en de verantwoordelijkheid van de zoekmachines. Strikt genomen zou dit betekenen dat burgers in veel gevallen (ook) bij de zoekmachines kunnen aankloppen.

1 Inleiding

Er is momenteel veel te doen over de bescherming van persoonsgegevens op internet. Niet iedereen is blij met de vermelding van zijn of haar persoonsgegevens op het World Wide Web.[1] Zo'n vermelding kan vervelende gevolgen hebben voor de persoon in kwestie. En ook al kan de gedupeerde partij hier tegen optreden via de Wet bescherming persoonsgegevens (WBP), het Wetboek van Strafrecht (smaad en laster) of het Burgerlijk Wetboek (onrechtmatige daad), toch zal het leed vaak al zijn geschied. Een zeer recent voorbeeld is actrice en mannequin Paris Hilton wier telefoon werd gekraakt door een hacker. Op diverse websites stonden in een mum van tijd niet alleen al haar contactgegevens van honderden bekende en minder bekende personen, maar ook haar privé-foto's welke waren opgeslagen in haar mobiele telefoon.[2] Het gevolg was dat niet alleen zij, maar haar hele vrienden-, kennissen- en familiekring een nieuw e-mailadres en telefoonnummer moest aanvragen. Ook hackers lijden echter onder de negatieve informatie-impact die het internet kan hebben op de persoonlijke levenssfeer. Een ex-hacker die geregeld (ongewild) diverse vermeldingen met betrekking tot zijn hackkunsten behaalde op allerlei sites in binnen- en buitenland, zei onlangs het hackende leven vaarwel en begon voor zichzelf een keurig adviesbureau dat cliënten adviseert over beveiligingsvraagstukken. Zoeken via Google naar de diensten van dit bureau leverde echter in de zoekresultaten van de zoekmachine ook alle minder aantrekkelijke wapenfeiten van deze gewezen hacker op. Naar inschatting blijven die vermeldingen tot in lengte van dagen op het internet rondzwerven met alle gevolgen van dien. Deze inschatting wordt nog eens bevestigd door het feit dat er ook zoekmachines zijn zoals de internet archive's 'wayback machine' die sinds 1996 internetsites die ooit het daglicht zagen op internet, tot in lengte van dagen vastlegt voor het 'nageslacht'.[3] Het kan zijn dat u op een website ooit informatie (over andere personen) heeft gepubliceerd waar u later, om welke reden dan ook, spijt van heeft gekregen. Deze zoekmachine helpt anderen daar nog lang aan herinneren!

Niet alleen de bekende sterren, maar ook de achteloze internetgebruiker kan slachtoffer worden van ongewenste vermelding(en) op internet. Recentelijk stond het klantenbestand van de huizensite Funda open en bloot on line. Het klantenbestand en enige andere informatie was door Google geïndexeerd waardoor men via Google niet alleen dit klantenbestand, maar ook inlogcodes voor de database en FTP-server kon vinden.[4] En wat ten slotte te denken van al die ex-vriendjes die de meest schandalige dingen, inclusief zeer onthullende foto's van hun ex-vriendin(nen), op internet plaatsen waardoor het 'Googlen' op de naam van de betreffende persoon[5] tot pijnlijke resultaten leidt met alle sociale gevolgen van

1 *Formeel gezien is het (klassieke) internet niet hetzelfde als het World Wide Web. Sinds het internet in 1992 werd opengesteld voor het grote publiek door lancering van het World Wide Web, heeft het bijna alle andere onderdelen van het internet overvleugeld. De termen internet en World Wide Web zullen dan ook door elkaar worden gebruikt zonder dat er een verschil in betekenis of reikwijdte aan moet worden verbonden.*

2 *Zie http://www.theregister.co.uk/2005/02/21/paris_hacked/.*

3 *Zie http://www.archive.org/.*

4 *Zie http://frontpage.fok.nl/nieuws/51487.*

5 *Bekendst voorbeeld is wel het gestolen dagboek van de Utrechtse studentes dat werd gescand en op internet gepubliceerd. Zie ook Rechtbank Amsterdam 7 april 2004.*

6 *Nu Google ook satellietbeelden gaat aanbieden kan men ook nog eens vervelend bezoek verwachten. Zie http://maps.google.com/ en http://story.news.yahoo.com/news?tmpl=story&cid=528&e=16&u=/ap/20050405/ap_on_hi_te/google_maps.*

dien.[6] En zo zijn er nog tientallen voorbeelden te noemen.

2 Indexeren van informatie: Google

Omdat Google één van de meest gebruikte zoekmachines ter wereld is en het woord 'Googlen' inmiddels als synoniem voor het woord zoeken wordt gebruikt, is het goed om in het bijzonder even bij Google stil te staan.[7] Een zoekmachine als Google struint hele dagen alle links op internet af en indexeert deze. Dit inclusief de andere pagina's die zij via een link op het bezochte domein tegenkomt. De inhoud van alle te indexeren pagina's wordt gelezen door de Googlebot of -spider.[8] De tekst wordt overgebracht naar een enorme database,[9] het andere deel van de zoekmachine.[10] De hyperlinks die de bot of spider vindt naar andere pagina's of andere sites worden gevolgd om ook die pagina's binnen te halen.[11] De kans dat daar sites en documenten met uw persoonsgegevens bij zitten die (eigenlijk) niet voor publicatie zijn bestemd is bijzonder groot.

Google houdt zich bij het indexeren en speuren over het web overigens netjes aan de The Digital Millennium Copyright Act (DMCA)[12] door bepaalde zoekresultaten, die mogelijk in strijd zijn met de DMCA, uit haar database te schrappen en hiervan melding te maken bij de zoekresultaten.[13] Dit veelal inclusief de vermelding van de DMCA-klacht.[14] De bescherming van de persoonlijke levenssfeer lijkt echter nog niet hoog op de agenda van Google en andere zoekmachines te staan. Het enige dat men in het privacyreglement van Google kan vinden over privacy met betrekking tot zoekresultaten is het volgende: 'De sites weergegeven als zoekresultaten of sites doorverwezen door Google Search Services zijn ontwikkeld door mensen waar Google geen invloed op uitoefent. Andere links, zoals voor het Google-vrienden mailing list archief, zijn ook sites die niet onder het beheer van Google vallen'.[15]

Na de lancering van haar Gmail[16]-dienst in 2004 waarin de gebruiker ermee akkoord moet gaan dat Google's bot gebruikers e-mailberichten leest opdat zij gericht kan adverteren, biedt ook niet veel hoop dat dit onderwerp meer prioriteit zal krijgen.

3 Vraag naar de verantwoordelijke

Veel mensen die menen dat hun persoonsgegevens ten onrechte op internet staan en daardoor te vinden zijn in de zoekresultaten van diverse zoekmachines, zoeken in de praktijk veelal hun toevlucht tot degene die deze erop heeft gezet, de hostingprovider, de rechter of het College bescherming persoonsgegevens. Voordat men toekomt aan de beoordeling en dus de toelaatbaarheid van de inhoud van diverse publicaties op internet, hetgeen raakt aan het spanningsveld tussen vrijheid van meningsuiting en de bescherming van de persoonlijke levenssfeer, is de eerste cruciale stap om vast te kunnen stellen wie eigenlijk verantwoordelijk is. Dit kan een zeer tijdrovende aangelegenheid zijn.
De verantwoordelijke geeft namelijk vaak niet thuis of is helemaal niet of bijzonder moeilijk te vinden.
Aankloppen bij de hostingprovider is ook niet zo makkelijk als het lijkt. De hostingprovider van de verantwoordelijke kan overal ter wereld zitten en zal zich liever ook niet met de inhoud bemoeien van hetgeen op de server staat (tenzij het een evident geval van racisme of kinderporno betreft).[17] Providers hebben nu eenmaal een commercieel belang bij het doorgeven van zoveel mogelijk informatie en willen absoluut niet in het nieuws komen als provider die vrij makkelijk informatie blokkeert. Het verkrijgen van de persoonsgegevens van de vermeende inbreukmaker die men wil aanspreken is vaak dan ook geen gemakkelijke klus. Denk hierbij ook aan de *Lycos*-affaire waarin een postzegelhandelaar, die met vermelding van naam, Pessers, en e-mailadres werd beschuldigd van fraude door een persoon die bij Lycos webruimte huurde, van Lycos verwijdering van de website en gegevens over de eigenaar vorderde. Lycos liet weten daar niet aan mee te willen werken, maar waarschuwde overigens de websitehouder wel voor de juridische consequenties die zijn actie met zich mee kon brengen.[18] Uiteindelijk werd Lycos in hoger beroep door de rechter veroordeeld tot het verstrekken van de naw-gegevens van de websitehouder. Veel had Pessers overigens niet aan deze uitspraak van het hof. De websitehouder had namelijk valse gegevens opgegeven aan Lycos toen deze destijds een webaccount bij Lycos nam. Pessers kon daardoor uiteindelijk nog niemand aanspreken.[19]

Wie besluit zich te wenden tot de toezichthouder (het CBP) of de rechter doet er in de regel verstandig aan om de klacht eerst bij de verantwoordelijke voor te leggen. In een gunstig scenario is het namelijk heel goed mogelijk dat de verantwoordelijke websitehouder niet opzettelijk de informatie op het web heeft gezet of dit niet zou hebben gedaan wanneer deze de precieze omstandigheden had gekend. In dat geval kan de verantwoordelijke alsnog de fout corrigeren en is ingrijpen van een derde partij helemaal niet nodig. Bij een minder gunstig scenario waarin de verantwoordelijke niet kan worden benaderd of waar het probleem zich voordoet dat de betreffende persoonsgegevens al op diverse andere sites zijn ondergebracht, geldt dat ook voor de toezichthouder het achterhalen van de (eerste) verantwoordelijke evenmin een gemakkelijke klus zal zijn. Een eventuele gang naar de rechter zou in een dergelijk scenario overigens nog gefrustreerd kunnen worden door het feit dat men wel eerst helder zal willen (en moeten) hebben wie men nu eigenlijk dagvaardt, mocht het zover komen. Hoewel de toezichthouder en de rechter (afhankelijk van de procedure) meer middelen ter beschikking staan dan de burger om de onderste steen boven te krijgen, moet hierbij

7 Er zijn overigens meer zoekmachines die van de Google-technologie gebruikmaken zoals A9. Verder is hetgeen in dit artikel wordt opgemerkt over Google ook van toepassing op veel andere zoekmachines op internet.

8 Bij het verzamelen en actualiseren van de gegevens worden veelal softwarehulpmiddelen ingezet die onder verschillende benamingen door het leven gaan. Zo zijn er de softwarerobots (bots), crawlers, spiders, enzovoort. Deze programma's laten zich allen scharen onder de noemer intelligent agents. De (Google) bots of spiders leiden voornamelijk een eigen leven. Iedere site waarnaar wordt gelinkt, zal per definitie ooit een keer worden gevonden. Wanneer een bepaalde pagina is verwijderd van een site zal deze nog steeds via de Google cache (de Google cache bevat kopieën van alle pagina's die zij indexeert) worden opgevraagd totdat deze na vaak lange tijd eindelijk verdwijnen.

9 Google beschikt met meer dan 3 miljard webpagina's over de grootste index ter wereld.

10 De kern van de zoektechnologie van Google heet PageRank. PageRank wijst aan elke pagina op het web een waarde voor de relevantie toe zodat kan worden vastgesteld hoe belangrijk de pagina is en wat (indirect) de hoogte zal zijn in de zoekresultaten.

11 Vindt Google bijvoorbeeld een link op site x naar een pagina op site y dan zal zij naast voornoemde link ook de andere aanwezige pagina's op site y indexeren.

12 The Digital Millennium Copyright Act Pub. L. No. 105-304, 112 Stat. 2860 (Oct. 28, 1998).

13 Zo'n vermelding kan er als volgt uitzien: 'Dankzij een klacht van Digital Millennium Copyright Act, hebben we 3 resultaten verwijderd van deze pagina. Indien gewenst, kun je

wel worden opgemerkt dat op het moment dat persoonsgegevens van iemand op internet verschijnen de klok gaat lopen. Het gevaar dat de gegevens verder worden verwerkt (door andere sites en zoekmachines) is alom aanwezig. Wanneer er ook nog eens sprake is van (mirror-)hosting van de betreffende gegevens in andere landen is er niet alleen de kans dat men door de bomen het bos niet meer ziet, maar neemt ook de kans toe dat men de ingezette race tegen de klok verliest. Voordat men het weet staat men tot in de eeuwigheid op internet. Toen columnist Luuk Koelman in 2003 van de rechter te horen kreeg[20] dat hij te ver was gegaan met zijn stukje over Gretta Duisenberg ('Menselijk schild in Ramallah') en de betreffende tekst van zijn site moest verwijderen, stonden de verwijderde passages binnen een uur na het vonnis op diverse mirror sites waaronder geenstijl.nl en retecool.com. Dit tot op de dag van vandaag.

Bijkomend probleem is dat men zich steeds vaker geconfronteerd ziet met diverse (mirror-) sites waaronder de webloggers waar de informatie razendsnel wordt hergebruikt en waarbij in de reacties van bezoekers bij de weblogs ook vaak nog het een en ander wordt opgediept over degene die op dat moment in de 'spotlight staat' (dit zijn tegenwoordig ook doorsnee burgers).[21] Voordat men het weet gaat de informatie pardoes een eigen leven leiden. Wie moet men dan nog aanspreken? En al kan men iemand aanspreken; in het slechtste geval moet men daarna nog tig mensen aanmanen om de informatie niet te vermelden casu quo te verwijderen. Daarbij kan het ook nog zo zijn dat de betreffende informatie (persoonsgegevens) van een bepaalde persoon op site A wel in de juiste context of binnen grenzen van toelaatbaarheid wordt gebruikt, terwijl op een andere site men net over het randje gaat (al dan niet door contextverandering of toevoeging van bepaalde informatie die de publicatie/vermelding (deels) ontoelaatbaar maakt). Dan hebben we het nog niet eens over regelgeving die per land kan verschillen. Het aangehaalde stukje over Gretta Duisenberg verwijderen van een Mexicaanse mirror is in beginsel dan ook onbegonnen werk. De vraag naar de daadwerkelijke verantwoordelijke laat zich overigens het moeilijkst beantwoorden bij subdomeinen[22] welke zijn verdisconteerd in de domeinnaam van de overkoepelende host. Bij dergelijke domeinnamen heeft men vaak te maken met websitehouders die bij het aanvragen van een webaccount om hun gegevens te publiceren valse/onjuiste informatie hebben opgegeven aan de hostingprovider.

Opvallend aan klachten of bemiddelingsverzoeken met betrekking tot ongewenste publicaties van persoonsgegevens op internet die binnenkomen bij de toezichthouder[23] is het feit dat bij de klacht of het verzoek tot bemiddeling bijna altijd de zoekmachine wordt genoemd via welke men de vermeende privacyinbreuk op het spoor is gekomen of via welke men de betreffende vermeldingen op het web kan vinden. Dat is eigenlijk ook niet zo vreemd omdat veel informatie alleen via een zoekmachine gevonden kan worden. Deze leidt de zoekende dan naar sites met de betreffende persoonsgegevens die deze persoon normaal wellicht niet zou hebben gevonden of bezocht. Mensen komen er via een zoekmachine vaak dan ook achter dat hun gegevens op internet staan. Een groot deel van de burgers zal dan ook niet zozeer problemen hebben met het feit dat (niet heel gevoelige) persoonsgegevens op een site voorkomen, maar wel met het feit dat men hiermee (voor anderen) mondiaal in het zoekresultaat van een zoekmachine is te vinden. Dat men op internet overigens voornamelijk graag naar informatie over andere mensen zoekt blijkt wel uit de statistieken van Google's Zeitgeist.[24]

4 Neveneffecten

De 'campagne' van de heer Zegers tegen Google[25] onderstreepte nog eens waar momenteel de schoen wringt toen voornoemde tegenstander van Google (echter wel strijder voor free weblogs) zich bij de Tweede Kamer, omroepen en nieuwsbladen beklaagde over het feit dat gegevens over hem te vinden waren in Google over reacties die hij onder andere in een forum op een omroepsite had achtergelaten. Terstond werd een item in geenstijl.nl aan hem gewijd waarin niet alleen de draak werd gestoken met zijn bezwaren,[26] maar waarin tevens alle links die leiden naar persoonsgegevens van de heer Zegers op internet waren vermeld. Iets dat hij nu juist niet wilde. Hiermee ontstond een beeld van de heer Zegers die al zijn zakelijke en amoureuze privé-activiteiten op internet bestreek. De gegevens waren op zichzelf staand redelijk onschuldig, maar de bijeengebrachte combinatie hiervan zorgde voor een negatieve beeldvorming jegens deze persoon. De links die de forumgasten over Zegers posten op geenstijl.nl met gegevens over Zegers waren ook hoofdzakelijk verkregen via Google.
Het feit dat Google nu juist alle links afstruint op internet en standaard een hogere ranking toekent aan hetgeen in forums is gepost (wat de vindbaarheid van die informatie doet vergroten) maakt het probleem duidelijk.

Ander bijkomend probleem is dat door al dat gespeur door zoekmachines naar nieuwe te indexeren[27] informatie, steeds meer privacygevoelige informatie die bij voorbaat niet eens bedoeld kan zijn om publiekelijk te worden gemaakt, in de openbaarheid komt. Hierbij moet men denken aan wachtwoorden, creditcard- en sofi-nummers en vertrouwelijke documenten die opeens via zoekmachines zijn op te vragen.[28] Wie zijn of haar wachtwoord of creditcardnummer intikt als zoekterm, loopt kans om op dergelijke onbeveiligde documenten te stuiten. Doordat Google sinds kort niet alleen PDF-bestanden, maar ook Powerpoint-, Word-, Excel- en RTF-bestanden doorzoekt is het risico alleen maar toegenomen.

de DMCA-klacht lezen voor deze verwijderde resultaten.' Zie ook http://www.google.nl/search?hl=nl&q=kazaa+lite&meta=.

14 Zie http://www.google.com/dmca.html. Zie ook http://www.copyright.gov/legislation/dmca.pdf.

15 Zie http://www.google.nl/intl/nl/privacy.html.

16 Google biedt via Gmail een gratis e-mailaccount met een opslagruimte van 2 gigabyte. In ruil daarvoor leest de Googlebot alle e-mails en bepaalt welke reclameboodschap zij gebruikers voorschotelt. Koppeling met het zoekgedrag is nog steeds niet als mogelijkheid uitgesloten.

17 Zie ook: W.Ph. Stol, Kinderporno op Internet, Nederlandse Politie Academie.

18 Hof Amsterdam 24 juni 2004 (Lycos/Pessers).

19 Voor meer informatie over verhouding tussen websitehouder en provider en aansprakelijkheidskwesties zie D. Lakerveld, Verboden links. Kan een hyperlink inbreuk maken op rechten van anderen?, scriptie privaatrechtelijke afstudeerrichting, december 2002. Zie http://www.uu.nl/content/Verbodenlinks.pdf.

20 Vzr. Rb. Breda, 125987/KG ZA 03-576. Het Hof Amsterdam heeft begin dit jaar het vonnis overigens vernietigd.

21 Denk aan het 'kijkerskanon' domme Debbie op geenstijl.nl.

22 Bijvoorbeeld pietsweb.myhost.net.

23 Zie ook p. 30 en 49 jaarverslag van het CBP 2004.

24 Zie http://www.google.com/press/zeitgeist.html.

25 Google is overigens niet de enige grote zoekcomputer. Inktomi is een andere grote speler welke de zoektechnologie van HotBot, AOL, iWon en Yahoo verzorgt.

En alsof dit allemaal nog niet lastig genoeg is, heeft de on line gemeenschap onlangs een nieuw zoekgezelschap mogen verwelkomen onder de naam ZoomInfo. Deze nieuwe zoekmachine richt zich in zijn geheel op persoonsgebonden informatie. Waar het bij Google nog enigszins verdedigbaar zou zijn om in veel gevallen van bijvangst te spreken wanneer een zoekopdracht (onbedoeld) eigenlijk niet voor publicatie (of niet voor publicatie in combinatie met andere gevonden gegevens) bestemde persoonsgegevens oplevert; bij ZoomInfo streeft men ernaar een zo compleet mogelijk beeld te kunnen bieden over de gezochte persoon.

ZoomInfo is net als Google een Amerikaans initiatief.[29] Via deze zoekmachine kan worden gezocht naar persoonsgegevens van meer dan 25 miljoen mensen en dat is nog maar het begin. Het verschil tussen deze zoekmachine en Google is dat bij ZoomInfo na het verzamelen van de bronnen die over personen op internet staan deze ook nog eens worden gerangschikt naar persoon en werkgever. De zoekmachine ZoomInfo legt, indien deze bestaan, ook nog eens verbanden tussen de verschillende personen waarover zij informatie heeft gevonden.

Daar waar persoonsgegevens in sterk gefragmenteerde vorm op internet staan en verder niet zoveel kwaad kunnen, zorgt Google dat dit gefragmenteerde plaatje in een meer geordend verband wordt gebracht. In beginsel niets zeggende gegevens verworden door ordening en koppeling van deze gegevens tot persoonsgegevens en krijgen daardoor betekenis. ZoomInfo gaat echter nog een stap verder en komt enigszins in de spionagesfeer terecht door het als resultaat van de zoekopdracht complete profielen met persoonsgegevens van de gezochte personen te bieden. Met zoekmachines als ZoomInfo hoeft een baanzoekende eigenlijk geen cv meer te sturen bij zijn sollicitatie. De aankomende werkgever kijkt gewoon even in ZoomInfo. Mocht de sollicitant wel een cv hebben ingesloten dan kan de P&O-afdeling dat nog eens rustig verifiëren via ZoomInfo.

Ondanks het gebruik van een combinatie van verfijnde kunstmatige intelligentie en natuurlijke taaltechnieken blijft ZoomInfo echter nog steeds de toepassing van machinale algoritmen[30] welke geen menselijke bemoeienis of enige vorm van redactie kent. Probleem van ZoomInfo is dan ook dat ondanks de soms zeer gedetailleerde profielen die het kan ophoesten ook records voorkomen die fouten bevatten. Zo was volgens ZoomInfo J. Battelle de directeur van de Northen Light zoekmachine terwijl deze persoon daar helemaal niets mee te maken heeft. Het gevaar hiervan is weer dat geboden profielen dus ook nog eens onjuiste persoonsgegevens kunnen bevatten, hetgeen voor de betrokken persoon in kwestie nog verdergaande gevolgen kan hebben dan een onrechtmatige c.q. ongewenste, doch juiste vermelding in de zoekresultaten. ZoomInfo vermeldt namelijk nergens dat zoekresultaten niet per se geheel juist hoeven te zijn. Eigenlijk geldt hetzelfde bij Google. Dat wat Google over u ophoest heeft u vaak niet eens kunnen controleren en hoeft ook niet eens te kloppen. Het feit dat Google sinds kort via maps.google.com zeer gedetailleerde satellietbeelden aanbiedt van steden en wijken is naast de dienst van ZoomInfo een ander voyeuristisch hoogtepunt hetgeen tevens duidelijk maakt dat de ambities van Google voor een deel aardig in de richting gaan van ZoomInfo. Dat dit problemen kan opleveren moge duidelijk zijn. Op de site www.michelvanrijn.nl kan men bijvoorbeeld vinden wie welke kostbare schilderijen aan de muur heeft hangen. Via Google maps is het dan handig zoeken naar waar dit kunstwerk precies hangt en wat de beste vluchtroute is. De zoekresultaten bevatten namelijk een prima overzicht.[31] Saillant detail is dan ook de enorme zoomslider die de beelden zo dicht op je netvlies kan brengen dat je het schilderij haast in de huiskamer kan zien hangen. Naar verluidt is het streven van Google om niet alleen foto's die (van bovenaf) door de satelliet zijn gemaakt op te nemen in de zoekresultaten, maar ook foto's die in een straat zelf zijn genomen. Google blijft overigens niet de enige. MSN (Microsoft) gaat nog wat verder en zal deze zomer haar eigen variant lanceren onder de naam MSN Virtual Earth. De beelden die Microsoft levert zijn onder 45 verschillende hoeken genomen. De gebruiker heeft hierdoor niet alleen toegang tot foto's die loodrecht op de oppervlakte genomen zijn.[32]

Sommige van de hiervoor genoemde aspecten doen een beetje denken aan Kafkaiaanse toestanden, maar dan op het internet. Er vindt door zoekmachines een compleet ander gebruik van persoonsgebonden informatie plaats dan waar deze oorspronkelijk voor bedoeld was. Velen herinneren zich ongetwijfeld de legendarische passage in het boek 'der Prozess'[33] van Kafka: 'Iemand moest Josef K. belasterd hebben, want zonder dat hij iets kwaads gedaan had, werd hij op een ochtend gearresteerd.' Nadat de hoofdpersoon Josef K. in het boek 'der Prozess' is gearresteerd wordt hij echter niet opgesloten, maar kan hij gewoon weer aan het werk gaan. Wel moet hij zich op gezette tijden melden voor de verhoren. Hoewel Josef K. niet fysiek gevangen zit, speelt de gevangenschap zich voornamelijk in zijn hoofd af. Steeds wanhopiger vraagt hij zich af wie verantwoordelijk is en hem antwoord kan geven op de vraag wat hij eigenlijk heeft misdaan dat hem dit overkomt. Alhoewel dit verhaal zich in een ver verleden afspeelt is bij gebruik van informatie die gevolgen kan hebben voor anderen de vraag naar de verantwoordelijke een onmiskenbaar stuk van de puzzel. Josef K. accepteert echter gaandeweg zijn lot om er uiteindelijk maar in te berusten. Ondanks het privacybewustzijn dat via de Privacyrichtlijn[34] in nationale wetgeving van de lidstaten is beland is voor velen de positie van voornoemde zoekmachines een gegeven waarmee je nu eenmaal maar moet leven. Velen zien het niet eens als een probleem. Een jonge moeder die 's ochtends

26 Mijns inziens werd er door geenstijl.nl ten onrechte aan voorbijgegaan dat een groot deel van de internetgebruikers in Nederland niet weet dat wanneer men iets post op internet, dit ook via zoekmachines kan worden gevonden. Mensen worden hier ook niet over voorgelicht op websites die men bezoekt.

27 De software waarmee een zoekmachine pagina's indexeert, wordt een robot, spider of crawler genoemd. De robots van de verschillende zoekmachines hebben een eigen naam, bijvoorbeeld: Googlebot (Google), Ingrid (Ilse), Scooter (Altavista). Bij het indexeren worden pagina's opgeslagen in de database van de zoekcomputer.

28 Opgemerkt dient te worden dat hier ook wordt gedoeld op zoeksystemen die men op de eigen computer kan installeren en welke vaak nog meer kunnen ontsluiten dan de grote zoekmachines.

29 Zie http://www.zoominfo.com/.

30 Een algoritme is een voorschrift, uit één of meer stappen opgebouwd, met al de handelingen die men achtereenvolgens moet verrichten om vanuit een gegeven beginsituatie tot het gewenste eindresultaat te komen. Dit kan tot uitvoering worden gebracht via een programma. Elke zoekmachine gebruikt overigens een ander algoritme om resultaten en de volgorde hiervan te bepalen. Van belang zijn bijna altijd: zoekterm in de URL, zoekterm in de titel, zoekterm in de omschrijving, zoekterm komt vaker voor in de tekst en linkpopulariteit. Over algoritmen zie ook W. De Raedt e.a., Algoritmen in, Inform, Basisboek 4.1, Kapellen: Uitgeverij Pelckmans 1998, p. 72-84.

31 Deze site was in het verleden bereikbaar onder http://www.michelvanrijn.com, maar verdween na een juridische aanvaring met gasmagnaat James Farrell op

wordt opgebeld door haar beste vriendin met de schokkende mededeling dat men via Google kan vinden dat zij (de jonge moeder) een 'hoer' is die haar kind mishandelt zal hier echter anders over denken. Naaste familie zal gauw begrijpen dat het om een wraakactie van de ex-man van de jonge moeder gaat en dat de informatie dan ook niet juist is. Voor de rest van de wereld, de Googlebot en ZoomInfo (wat een afschuwelijk profiel zal dat opleveren) is dat wat minder transparant.

5 Omgekeerd zoeken

Daar waar Google en aanverwante zoekmachines nog enige gaten laten liggen worden die opgevuld door de zogenaamde 'omgekeerd zoeken' zoekmachines. Hierbij kan men denken aan sites zoals gebeld.nl, ikhebje.nl, zoekenbel.nl, qik.nl, zoekopnummer.nl en spyderweb.nl. Dergelijke internetbedrijven bieden diensten aan die vroeger alleen door de traditionele telecommunicatiebedrijven werden geleverd, maar met (omgekeerde) zoekmogelijkheden die de originele gidsen en directories niet kennen. Dat de traditionele (papieren) abonneelijsten op het internet beschikbaar komen is een normale zaak. De bekende Telefoongids kan men immers ook via internet raadplegen. De 'omgekeerd zoeken' sites voorzien deze traditionele producten echter van allerlei nieuwe zoekcriteria, welke zoveel functionaliteit toevoegen aan het oorspronkelijke product of dienst dat er nog moeilijk van eenzelfde dienst kan worden gesproken.

Was het bij de oorspronkelijk gidsen nog vrij helder wat dit betekende voor de privacy van de betrokkenen; de huidige producten/diensten zijn voor de betrokkenen nog weinig transparant. Uit eenzelfde set van gegevens kan ineens veel meer informatie worden gehaald dan de betrokkene denkt. Zeker wanneer men bedenkt dat een koppeling met andere openbare bestanden zoals stadsplattegronden tot een schat aan informatie over iemand kan leiden. Het verschil met een zoekmachine als Google is wel dat het domein van dergelijke zoekmachines veel beperkter is (alhoewel hier ook uitbreiding plaatsvindt). Veelal maken deze zoekmachines gebruik van oude klantendatabases, telefoongidsen en andere databases met naw-gegevens die vaak niet bestemd zijn voor 'people finding'.

Was het tot voor kort zo dat abonnees automatisch in een abonneelijst[35] werden opgenomen tenzij men daar bezwaar tegen maakte, sinds het van kracht worden van het gewijzigde art. 11.6 Telecommunicatiewet (Tw) dienen uitgevers van abonneelijsten (bijvoorbeeld telefoongidsen) toestemming te vragen aan de abonnee. Onder abonneelijst wordt niet alleen de telefoongids, maar ook elektronische of gedrukte gidsen verstaan waarin e-mail (en andere contactgegevens) van abonnees worden opgenomen mits deze algemeen beschikbaar zijn.[36] Verder rust op de uitgever van de abonneelijst of degene die de algemeen beschikbare abonnee-informatiedienst verzorgt, enige informatieverplichtingen. De abonnee dient op de hoogte te worden gebracht van de doeleinden van de abonneelijst of abonnee-informatiedienst en in het geval dat het om een elektronische versie van de abonneelijst gaat, van de gebruiksmogelijkheden op basis van daarin opgenomen zoekfuncties. Verder dient de abonnee geïnformeerd te worden over de soorten persoonsgegevens die daarin kunnen worden opgenomen. Opgemerkt dient te worden dat toestemming voor opname van persoonsgegevens op basis van art. 11.6 lid 2 Tw, kan worden gevraagd door de gidsuitgever of door een derde.[37]

In bepaalde (on line) elektronische gidsen kan men zoals we hiervoor reeds zagen dus ook 'omgekeerd zoeken', hetgeen inhoudt dat men een naam bij het telefoonnummer kan vinden. Bij de officiële elektronische uitgaven zoals de telefoongids kan dat niet. Voor een dergelijke mogelijkheid (lees: zoekfunctionaliteit) dient men apart toestemming te vragen van de betrokkene.[38] Dit betekent dat een flink aantal (hobby)sites op internet waar de 'omgekeerd zoeken' mogelijkheid al enige jaren wordt aangeboden, toestemming zullen moeten verkrijgen van de personen die in de daarvoor ingerichte database zijn opgenomen en nog zullen worden opgenomen. De abonnee heeft daarnaast ook nog het recht van correctie en verwijdering met betrekking tot de persoonsgegevens die in de algemeen beschikbare abonneelijst zijn opgenomen of worden gebruikt ten behoeve van een abonnee-informatiedienst. Geen enkel 'omgekeerd zoeken' site heeft echter toestemming van alle abonnees die zijn opgenomen in haar databases. Toen onlangs een 16-jarige scholier werd mishandeld na een oproep op voelspriet.nl doordat men via het 'omgekeerd zoeken' de naw-gegevens van deze scholier had achterhaald, werd nog eens duidelijk hoe belangrijk niet alleen de bescherming van de persoonlijke levenssfeer is, maar ook dat dienstverleners hun verantwoordelijkheid daarin nemen.

Het is dan ook op zijn minst een beetje vreemd dat de onlangs aangepaste Telecommunicatiewet strenge regels hanteert met betrekking tot opname in elektronische gidsen en directories (welke bij de 'omgekeerd zoeken' sites nog niet echt lijken te worden nageleefd), maar dat men via Google wel zonder problemen allerlei adresgegevens kan vinden zonder dat de zoekmachine zich daarbij aan bepaalde regels dient te houden. Bij de bespreking van de zoekmogelijkheden kwam reeds naar voren dat via zoekmachines nu juist ook combinaties van gegevens worden gevonden die normaal niet in die samenstelling zonder toestemming in een abonneelijst of op een 'omgekeerd zoeken' site zouden mogen worden aangeboden.

onverklaarbare wijze uit de DNS-adresstructuur zodat niemand de informatie onder die domeinnaam nog kon waarnemen. Een manier van iemand de mond snoeren die niet eenieder is gegeven.

32 Zie ook http://www.pcworld.com/resource/article/0,aid,120968,pg,1,RSS,RSS,00.asp.

33 F. Kafka, Het Proces (1925), uitgeverij Athenaeum-Polak & Van Gennep.

34 Richtlijn nr. 95/46/EG van het Europees Parlement en de Raad van 24 oktober 1995 betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens en betreffende het vrije verkeer van die gegevens, PbEG L 281/31.

35 Bij abonneelijsten moet men niet alleen denken aan de klassieke telefoongidsen, maar ook aan elektronische gidsen op cd-rom of internet en (whois) directories.

36 Kamerstukken II 2002/03, 28 851, nr. 3, p. 158.

37 Nota van toelichting, Besluit van 7 mei 2004, houdende regels met betrekking tot universele dienstverlening en eindgebruikersbelangen (Besluit universele dienstverlening en eindgebruikersbelangen), p. 25.

38 Vergelijk art. 11.6 lid 3 Tw. Die toestemming was overigens ook al vereist voor de wijziging van de Telecommunicatiewet. De gegevens worden namelijk voor een ander doel gebruikt dan waarvoor ze zijn verkregen.

6 Semantisch web: de toekomst

Wie vindt dat men nu al veel kan vinden via zoekmachines over andere personen zal straks helemaal versteld staan van wat men zoal aan persoonsgegevens kan vinden. Internet groeit langzaam toe naar een semantisch web.

Het overgrote deel van de informatie die nu op internet beschikbaar is, is opgemaakt in HTML-documenten. Alhoewel met het gros van de huidige HTML-pagina's al behoorlijke zoekresultaten kunnen worden bereikt is HTML met name geschikt om de informatie op het web te visualiseren. Het is dan ook veel minder geschikt om die informatie zodanig te beschrijven dat men software kan gebruiken om deze informatie te vinden of te interpreteren. In HTML kunnen dan ook geen contextuele of conceptuele gegevens worden opgenomen. Dat is dan ook de reden dat een computerprogramma op dit moment bijvoorbeeld niet kan vaststellen of getallen in een webdocument een bedrag, een datum of een leeftijd voorstellen. De bestaande zoekmachines die voor de internetgebruiker zoeken op trefwoord ontsluiten naast bruikbare informatie dan ook bijzonder veel irrelevante informatie. Die irrelevante informatie ontstaat doordat de zoekcomputer niet precies weet in welke context hij de zoekopdracht moet plaatsen en welke betekenis precies aan bepaalde trefwoorden moet worden toegekend. De internetgebruiker die een zoekopdracht geeft wordt daardoor overstelpt met duizenden hits waarvan er maar een paar echt relevant zijn. Er wordt daarom hard gewerkt aan zoekmachines die door gebruik van ontologieën (beschrijvingen van concepten en hun onderlinge relaties) pagina's kunnen vinden die syntactisch verschillende, maar semantisch gelijksoortige woorden bevatten. Het internet wordt dan één groot semantisch web.[39] In een semantisch web wordt de informatie geannoteerd door middel van metadata[40] (gegevens over gegevens) zodat het ook door machines leesbaar is en kunnen er automatisch relaties binnen en tussen documenten gelegd worden.[41] Documenten op het web worden dus veel informatiever. Een (experimentele) zoekmachine voor semantische webdocumenten is overigens Swoogle.[42]

Het hoeft denk ik geen betoog dat dit semantisch web de bescherming van persoonsgegevens nog meer onder druk zal zetten. Men zal via de toekomstige zoekmachines nog meer, nog sneller en nog preciezere informatie kunnen vinden over mensen. Anderzijds kan bovenstaande ontwikkeling mogelijk ook oplossingen bieden voor de bescherming van persoonsgegevens. Het wordt namelijk makkelijker om te definiëren of het in een document om persoonsgegevens gaat en of hier nu wel of niet rekening mee gehouden moet worden.

7 Verantwoordelijkheid van zoekmachines

Zoals reeds opgemerkt staan op internet veel gegevens die op zichzelf geen persoonsgegeven (behoeven te) zijn, maar in combinatie met andere gegevens dit wel kunnen worden. Anderzijds zijn veel persoonsgegevens zonder zoekmachine niet eens vindbaar en daardoor ook niet zo relevant.[43] Op zichzelf staande gegevens worden echter door de gepresenteerde combinatie in Google persoonsgegevens welke herleidbaar zijn naar één of meer natuurlijke personen.[44] Hierdoor ontstaat een nieuwe dimensie die er zonder zoekmachine niet zou zijn. Het verschil met zo'n 'omgekeerd zoeken' site is (ondanks het kleinere domein en de meer specifieke informatie) niet zo groot als men denkt. Beide bedienen zich namelijk van een database. De één is echter veel groter en meer divers dan de ander. Feitelijk wordt bij beide constructies een grote database doorzocht. De één is duidelijk aan regels gebonden en de ander niet, terwijl ze min of meer hetzelfde doen en vaak zelfs dezelfde informatie opleveren.

Men kan zich afvragen in hoeverre een zoekmachine ongevoelig kan zijn voor het doel waarmee bepaalde persoonsgegevens op internet worden gezet. Dat internet een verzameling links is en voor iedereen toegankelijk betekent niet dat daarmee bepaalde doelen geheel uit het oog moeten worden verloren. Iemand die een advertentie op Marktplaats zet, staat er niet bij stil dat deze gegevens ook meteen in Google te vinden zijn. Hierover wordt degene die het bericht plaatst ook niet geïnformeerd. Zou men dit wel weten en dus geïnformeerd zijn, dan zou men wellicht minder persoonsgegevens opgeven. Want laten we eerlijk zijn. Iemand die een advertentie opgeeft op Marktplaats bij radioapparatuur, richt zich alleen op bezoekers van Marktplaats die radioapparatuur aanbieden of zoeken en niet op elke willekeurige internetbezoeker die aan het Googlen is. Internet gebruiken om een advertentie te zetten op Marktplaats betekent nog niet dat iemand daarmee een algehele publicatie zoekt op het World Wide Web. Moeten nieuwsgierige 'people finders' nu echt weten dat iemand al twee keer bij Marktplaats een radio heeft aangeboden waarbij tevens de naw-gegevens en het mobiele nummer staan vermeld? En ook al wordt de advertentie verwijderd op Marktplaats dan nog is deze maandenlang via Google te vinden. De inhoud van de oude advertentie kan men dan nog opvragen via de Google cache. Google stelt dat zij de zoekresultaten niet handmatig kan wijzigen. Wanneer de kopie uiteindelijk uit het cachegeheugen is verwijderd door Google worden de titel en de URL van de pagina nog wel weergegeven voor de betreffende zoekopdrachten, totdat de (marktplaats)site weer door de robots van Google wordt bezocht. Men moet dus wachten tot de Googlebot opnieuw de site bezoekt. Deze bot zal dan constateren dat de informatie bij Marktplaats niet meer bestaat en de inhoud van de index bijwerken.

39 Volgens W3C (het consortium dat de webstandaarden vaststelt) is het semantisch web een constructie waarin gegevens op het web zodanig gedefinieerd en gelinkt worden dat zij door machines niet alleen gebruikt kunnen worden voor presentatiedoeleinden, maar ook voor automatisering, integratie en hergebruik van gegevens via diverse toepassingen. Zie ook http://www.w3.org/2001/sw/.

40 Zie ook http://www.niso.org/standards/resources/UnderstandingMetadata.pdf.

41 De twee belangrijkste technologieën voor het ontwikkelen van het semantisch web zijn op dit moment: de Extensible Markup Language (XML) en het Resource Description Framework (RDF). XML en RDF stellen bouwers van sites in staat om hun informatie semantisch te markeren c.q. informatie over gegevens op internet te beschrijven. Zie verder ook Tim Berners-Lee, James Hendler & Ora Lassila, 'The Semantic Web', Scientific American, Mei 2001.

42 Zie http://swoogle.umbc.edu/index.php.

43 Zie ook R.C. Winkelhorst & M. van der Linden-Smith, 'Persoonsgegevens op Internet; een (ver)melding waard?', NJB 2004, p. 627-631, Deventer: Kluwer 2004.

44 Vergelijk art. 1 sub a WBP.

Marktplaats geeft overigens aan dat zij er niets aan kan doen dat Google haar pagina's inclusief persoonsgegevens indexeert. Dit alles bevestigt nog eens de constatering dat het bijzonder moeilijk is voor betrokkenen om de juiste verantwoordelijke te vinden en zijn of haar rechten te effectueren. Zelfs in het meest gunstige geval dat verantwoordelijken duidelijk zijn en dat het verzoek tot verwijdering wordt gehonoreerd, moet men nog maanden geduld hebben voordat de gegevens geheel zijn verwijderd uit de index van de zoekmachine. En ja, dan kan bepaalde informatie inmiddels alweer op een andere site of zoekmachine te vinden zijn en kan men weer van voren af aan beginnen. Wanneer betrokkene met een situatie te maken krijgt waarbij hij twee of meer verantwoordelijken (bijvoorbeeld de zoekmachine en de website waar de informatie op staat) nodig heeft om de informatie verwijderd te krijgen (iets staat op een forum en kan niet worden verwijderd door de betrokkene zelf), dan is alleen de medewerking van slechts één verantwoordelijke niet voldoende en is afstemming wenselijk.

Het argument van bijvoorbeeld Google dat webbeheerders zelf verantwoordelijk zijn voor de inhoud van hun sites en dat zij slechts links daartoe aanbieden kan om nog een reden niet opgaan. Google zou in die filosofie namelijk dan ook niet die sites uit haar index moeten verwijderen die mogelijk de DMCA overtreden. Kennelijk speelt hier het feit mee dat Google een Amerikaans bedrijf is en in Amerika gaat men anders om met bescherming van persoonsgegevens dan in Europa. Bovendien is de aansprakelijkheid onder de DMCA niet gering. In Amerika werd overigens de DCMA door de Scientology Kerk misbruikt om persoonsgegevens van een haar onwelgevallige Nederlandse website te verwijderen. Doordat in Nederland de DMCA niet op een Nederlandse provider van toepassing is, maar wel op een Amerikaanse zoekmachine, werd via een omweg toch hetzelfde resultaat bereikt door Google min of meer te dwingen de bewuste pagina's uit de index van de zoekmachine te verwijderen.[45] Hierdoor was de DMCA materieel gezien dus gewoon van toepassing op Nederland.[46]

Een mogelijk argument dat Google zich als Amerikaans bedrijf niet aan Europese privacyregelgeving hoeft te houden is niet zo sterk. Google verwerkt (ook) gegevens van Europeanen en zet haar zoekportaal in (en biedt dus haar diensten aan) onder de vlag van diverse nationale domeinen in Europa. Daarnaast heeft Google diverse verkoopkantoren en servers die informatie verzamelen en beschikbaar stellen (en dus niet alleen zijn bestemd voor doorgifte) opgesteld in diverse Europese lidstaten waardoor men ervan uit zou mogen gaan dat het nationale recht van de lidstaten op haar van toepassing is op grond van art. 4 lid 1 sub *c* Richtlijn nr. 95/46/EG (en voor Nederland art. 4 lid 2 WBP).[47]

Er zullen natuurlijk ook mensen zijn die zeggen: 'sites moeten zelf maar zorgen dat Google ze niet kan vinden'. En inderdaad: via een simpel tekstbestandje genaamd robots.txt kan men (in de root van de webdirectory) aangeven aan de bezoekende bot of spider wat hij wel en niet moet indexeren. Echter, om dit te kunnen aangeven moet men vrij veel rechten (en kennis) hebben met betrekking tot de gegevens welke op het domein worden gehost. Veel mensen die informatie op internet zetten hebben die mogelijkheden en kennis niet. De genoemde oplossing is echter niet 100% waterdicht. Anderzijds speelt ook een rol dat er ook weer zoekmachines zijn die juist de sites indexeren die in het robots.txt bestand zijn gemarkeerd als niet-indexeerbaar. Het gaat hier met name om e-mail harvesters, downloading agents, spam bots en leechware.[48]

Nu het zo moeilijk blijkt in de praktijk om de rechten te effectueren omdat in eerste instantie de partijen betrokken bij de informatiehuishouding zich allemaal (niet ten kwade overigens) achter elkaar verschuilen (de zoekmachine verwijst naar de websitehouder en de websitehouder verwijst weer naar Google waarbij moet worden opgemerkt dat veelal niet eens duidelijk is wie een bepaalde site houdt casu quo wie verantwoordelijk is voor bepaalde informatie op een domein), kan men zich dan ook afvragen in hoeverre er eigenlijk geen sprake zou moeten zijn van wat meer verantwoordelijkheid van de zoekmachine.

Een zoekmachine heeft de macht om sturing aan te brengen in wat wel en niet kan worden gevonden op internet en is een vrij centraal aanspreekpunt in de decentrale wirwar van sites en verantwoordelijken. Het lijkt dan ook minder praktisch de gehele oplossing voor het geschetste probleem neer te leggen op het niveau van de websitehouders. Deze groep is namelijk veel te divers en voor de burgers veelal een onbruikbaar aanspreekpunt. Denk hierbij aan hetgeen werd opgemerkt in het voorgaande over het verschil in kennis, gebruikte faciliteiten, doelen en (kwade) intenties. Verder zagen we in het voorgaande al dat in het zoekresultaat dat een zoekmachine ophoest een combinatie van informatie bijeen wordt gebracht waarvan de informatieonderdelen op zichzelf geen persoonsgegevens hoeven te zijn, maar door de gepresenteerde combinatie dit ineens wel kunnen zijn. Elke zoekopdracht kan dus in feite een nieuwe verwerking van persoonsgegevens opleveren. Hierbij dient men ook te denken aan die situaties waarin de burger met alleen de vermelding van die gegevens op een afzonderlijke website geen problemen heeft, maar met de vindbaarheid hiervan in Google wel. Daarbij komt dat Google en andere zoekmachines momenteel allerlei sites indexeren waarvan ze niet weten of deze documenten inclusief bepaalde persoonsgegevens die daarin staan vermeld zich wel lenen voor algehele openbaarmaking. In dat licht bezien zou het dan ook niet zo gek zijn wanneer een zoekmachine

45 Zie ook http://www.xtdnet.nl/paul/PrioriyTelecom-Xenu.html.

46 De US Children's Online Privacy Protection Act 1998 (COPPA) is overigens wel gewoon van toepassing op buitenlandse websites die persoonlijke informatie van kinderen op het grondgebied van de VS verzamelen. Volgens deze wet moet de beheerder van een website de COPPA naleven.

47 Zie ook Artikel 29 Werkgroep, 5035/01/NL/def. WP 56.

48 Dit zijn de zogenaamde 'bad' agents, zoals Whatsnew, Boitho, Imagefetch, Personapilot, Emailcollector, enzovoort.

meer verantwoordelijkheid neemt. Hierbij doel ik niet op een inhoudelijke toetsing van de informatie omdat dit op gespannen voet komt te staan met de vrijheid van meningsuiting, maar op een iets andere inrichting van het verzamelen van informatie.

Dit zou kunnen worden bereikt wanneer de bots, de spiders en crawlers stoppen met het (geautomatiseerd) beslissen wat wel en niet wordt geïndexeerd en dat websitehouders zelf de pagina's aanmelden bij de zoekmachine welke voor publicatie zijn bestemd (net als vroeger dus).[49] Bij de aanmelding zou de zoekmachine dan tevens moeten verlangen dat de verantwoordelijke websitehouder tevens contactinformatie (ten minste een e-mailadres) opgeeft zodat in geval van problemen contact kan worden gezocht met de juiste verantwoordelijke. Hiermee wordt voorkomen dat de betrokkene die zich met een onrechtmatige, onjuiste of ongewenste publicatie van zijn persoonsgegevens ziet geconfronteerd op internet niet weet wie hij of zij hiervoor moet aanspreken. Daarna kan (indien partijen er onderling niet uitkomen) door de rechter of toezichthouder worden bezien of de betreffende publicatie inderdaad onrechtmatig is of niet. Hiermee blijft de vrijheid van meningsuiting geheel intact.

Bij het zorgvuldig verwerken van persoonsgegevens hoort een verantwoordelijke welke aanspreekbaar dient te zijn op de verwerking van persoonsgegevens.[50] Indien een verantwoordelijke valse casu quo onjuiste gegevens opgeeft bij het aanmelden van zijn webpagina's dan zal deze persoon het voor lief moeten nemen dat de link uit de index van de zoekmachine wordt verwijderd (dit gebeurt nu overigens ook al bij ranking fraude en bij DMCA-inbreuken) indien betrokkene op terechte gronden over de publicatie klaagt en verantwoordelijke gewoonweg onbereikbaar is. Het gaat dan kennelijk om een (onrechtmatige) publicatie waarvoor verantwoordelijke geen enkele verantwoordelijkheid wenst te nemen. Vergelijk ook de overwegingen (met name r.o. 4.10) van het hof in de *Lycos/Pessers*-zaak waarin het hof het anoniem beschuldigen aan banden legt.[51] Hierbij moet overigens nog wel eens goed worden nagedacht hoe men dit het beste kan doen zonder de vrijheid van meningsuiting (te veel) onder druk te zetten. Anders dan in de *Lycos/Pessers*-zaak blijft in de hiervoor beschreven handelwijze de publicatie op de website zelf nog wel in stand.

Opgemerkt dient nog te worden dat naast een wijziging in de manier van vergaren van informatie door zoekmachines, een zoekmachine als Google ook het proces eigenlijk zo zal dienen in te richten dat bij een gehonoreerd verwijderingsverzoek de betreffende informatie ook binnen redelijke termijn[52] uit de zoekmachine verdwijnt en niet zoals nu het geval is, nog maanden op internet beschikbaar blijft.[53]

8 Conclusie: A bot is hot; privacy is not?

Internet wordt meer en meer bestuurd door bots, spiders en crawlers. Alhoewel deze allemaal kunnen worden geschaard tot de zogenaamde intelligent agents, zijn deze agents allesbehalve intelligent als het op het gebied van bescherming van persoonsgegevens aankomt. Het zijn echter dezelfde bots die momenteel wikken en wegen over onze persoonsgegevens. Zij brengen onze informatie bijeen in rap tempo en beslissen hoe we kunnen worden gevonden. Daar achteraan hollen de betrokkenen die zich via de zoekmachines steeds vaker geconfronteerd zien met onrechtmatige casu quo ongewilde publicatie van hun persoonsgegevens op internet. Verantwoordelijken verschuilen zich achter elkaar. Zoekmachines wijzen met de vinger naar de websitehouders welke op hun beurt weer naar de zoekmachines wijzen. Bij veel inbreuken kan overigens helemaal nergens naar worden verwezen omdat men niet weet wie verantwoordelijk is. Dit is vaak het geval bij kwaadwillende vermelding van persoonsgegevens op internet. Ondertussen heeft de betrokkene voortdurend het nakijken. Mocht het de betrokkenen lukken om de juiste verantwoordelijke(n) te vinden, dan duurt het nog vaak maanden voordat de informatie uit de index en/of de cache van de zoekcomputers is verdwenen.

Het belangrijkste argument om meer verantwoordelijkheid bij zoekcomputers te leggen daar waar het gaat om bij problemen een verantwoordelijke te kunnen aanwijzen, is het feit dat veel gegevens die op internet staan op zichzelf geen persoonsgegeven (behoeven te) zijn,[54] maar in combinatie met andere gegevens dit wel kunnen worden. Veel persoonsgegevens zijn zonder zoekmachine bovendien niet eens vindbaar en daardoor ook niet zo relevant. Op zichzelf staande gegevens worden echter door de gepresenteerde combinatie in de zoekresultaten van Google persoonsgegevens welke herleidbaar zijn naar één of meer natuurlijke personen.[55] Hierdoor ontstaat een nieuwe dimensie die er zonder zoekmachine niet zou zijn geweest. Daarnaast is Google in feite niet anders dan de 'omgekeerd zoeken' sites terwijl op laatstgenoemde groep wel specifieke regelgeving (de Telecommunicatiewet) van toepassing is. Bovendien is het vinden van een naam bij een creditcardnummer toch iets ernstiger dan het vinden van een naam bij een telefoonnummer. Gezien hetgeen in het voorgaande reeds werd opgemerkt met betrekking tot de manier waarop Google in Europa opereert en de toepasselijkheid van art. 4 lid 1 sub *c* Richtlijn nr. 95/46/EG, ligt het voor de hand dat zij zich ook aan bestaande regelgeving dient te houden en is de burger mogelijk niet zo vogelvrij als tot nu toe werd aangenomen.

Er zullen echter ook mensen zijn die vinden dat internet een publieke plek is en wanneer je wilt dat iets niet op

49 Om het efficiënt te houden zou men op een constructie kunnen overgaan waarin de websitehouder in één keer kan aangeven wat wel en niet voor publicatie is bestemd en mocht men alles willen aanmelden, dan geeft men de gehele URL op. In die constructie bezoekt de zoekmachine op het moment van aanmelding de opgegeven site en biedt een overzicht (let op: alleen een overzicht!) van de aangetroffen documenten op die site. Hierbij wordt de aanmelder dan een keuze gegeven middels een aanvinkscherm welke pagina's geïndexeerd moeten worden en welke niet.

50 Daar er vanuit de Verenigde Staten steeds meer geluiden opgaan voor mogelijk een (Intellectueel) Eigendomsrecht op persoonsgegevens is indirect een aardige link met de DMCA gelegd. Zie ook http://www.oreillynet.com/pub/wlg/2480.

51 Pessers kreeg door de uitspraak uiteindelijk wel de naw-gegevens van de inbreukmaker, maar die bleken vals.

52 Vergelijk ook art. 6, 7 jo. 10 WBP.

53 Dit geldt overigens ook voor MSN en de op de Inktomi gebaseerde zoekmachines.

54 Of persoonsgegevens betreffen van minder gevoelige aard.

55 Vergelijk art. 1 sub a WBP.

internet komt je er niets op moet zetten. In die redenering is het ook zo dat wanneer je niet wilt worden beroofd, je niet je huis moet verlaten en wanneer je niet wilt worden lastiggevallen door telemarketeers of spam, je respectievelijk geen telefoon of e-mailaccount moet nemen. Verder gaan deze mensen eraan voorbij dat veel informatie ook zonder toedoen van de betrokkene op internet terechtkomt. Het is niet onbelangrijk dat mensen toch een keus hebben of iets wel of niet voor iedereen beschikbaar is en dat transparant moet zijn wie iets op internet heeft gezet en dat dit ook snel weer kan worden verwijderd indien daar gegronde redenen aan ten grondslag liggen.

Met de komst van het semantisch web zullen de zoekmogelijkheden en daarmee de ontsluitbaarheid van het World Wide Web alleen maar toenemen en zal er een Wild West ontstaan waar uiteindelijk vanuit privacyoogpunt niemand bij gebaat is. Het zou zinnig zijn wanneer alle betrokken partijen eens goed over de mogelijke gevolgen en over hun rol daarin zouden nadenken.