

Herbert Braun, Holger Bleich, Mariska van Dasselaar, Noud van Kruysbergen

Webstatistieken

Methoden om websitegebruik te analyseren

Als je behoorlijk wat tijd en energie in je website steekt, is het ook een must dat je het websiteverkeer analyseert. Daarbij is het natuurlijk belangrijk dat je de privacygrenzen respecteert. Is het analyseren van websitebezoekers wettelijk eigenlijk wel toegestaan?



Om je website te verbeteren, kun je het beste beginnen met het in kaart brengen van de acties van je bezoekers. Zo kan een kleine ondernemer uit de logbestanden achterhalen via welke zoektermen de bezoekers bij zijn bedrijf uitkomen, kan de eigenaar van een webwinkel onderzoeken waarom potentiële klanten hun winkelwagentje wel vullen maar de site uiteindelijk verlaten zonder daadwerkelijke aankoop, en kan een blogger achterhalen welke onderwerpen niet alleen hem, maar ook zijn lezers interesseren.

Maar er zijn meerdere redenen waarom een bezoekersanalyse van je website van belang is. Dergelijke informatie kan namelijk ook gebruikt worden om reclame efficiënt in te zetten, afgestemd op de interesses van de bezoeker.

Het doel van bezoekersanalyse is dan ook om zoveel mogelijk over de surfer te weten te komen.

Volgnummer

'U bent bezoeker nummer 1717'. In de prehistorie van het internet behoorden kleine CGI-scripts voor het tellen van websitebezoekers tot het standaardrepertoire van een website. Tegenwoordig kom je die alleen nog tegen op kleine hobbysites en bij gratis websitebouwpakketten. Een webmaster valt zijn bezoekers liever niet meer lastig met de resultaten van een telscript en heeft een discrete statistische analyse op de webserver van daar verzamelde gegevens de voorkeur.

Bij de aanroep van een website stuurt de browser een heel

pakket aan gegevens naar de webserver. Firefox vraagt bijvoorbeeld met de volgende header de c't-website op:

```
GET /ct/ HTTP/1.1
Host: www.fn.nl
User-Agent: Mozilla/5.0 (Windows; U;
          Windows NT 5.1; nl; rv:1.9.0.5)
Gecko/2008122203 Firefox/3.0.5
Accept: text/html,application/
       xhtml+xml,application/xml;
       q=0.9,*/*;q=0.8
Accept-Language: nl,en-us;q=0.7,en;q=0.3
Accept-Encoding: gzip,deflate
Accept-Charset: ISO-8859-1,utf-8;
               q=0.7,*/*;q=0.7
Keep-Alive: 300
Connection: keep-alive
Referer:http://www.fn.nl
```

Hiermee meldt de client wat de gevraagde webpagina is en wordt er informatie gegeven over

de verbinding. Voor de dataoverdracht kunnen hier de gangbare MIME-typen voor HTML en XML en transport- en tekstcoderingen gebruikt worden. De host komt daarbij nog een paar interessante dingen te weten, bijvoorbeeld dat de client zich Firefox 3.0.5 noemt en dat die volgens deze informatie op een computer met Windows XP (alias 'NT 5.1') draait. De gebruiker van deze computer komt uit Nederland, maar accepteert ook Engelstalige teksten.

Je kunt er overigens niet zondermeer vanuit gaan dat al deze gegevens ook waar zijn. Er bestaat namelijk altijd een mogelijkheid dat deze verkeerd zijn geconfigureerd. Toch kloppen ze in de praktijk vaak wel. Overigens bevat de laatste regel belangrijke informatie voor een webstatisticus: de referrer (in HTML 'referrer'

omdat die vaak verkeerd werd gespeld) is de URL van waaruit de gebruiker op het nu aangevraagde adres is terechtgekomen, bijvoorbeeld doordat hij daar een link heeft aangeklikt. Als je een URL rechtstreeks in de browser intypt, is er dus geen referrer.

Bij deze HTTP-gegevens komt dan nog het via TCP/IP verstuurd afzenderadres. Deze vier bytes lokaliseren niet alleen het adres van de computer op internet, maar ook de geografische locatie van de gebruiker. Een dienst als ip2location.com laat naast het IP-adres van je computer meteen de stad, de tijdzone en de lengte- en breedtegraad zien. Al moet je dat wel met een korreltje zout nemen, omdat dat dan meestal van de server van je provider is en het niet je eigen geografische gegevens zijn.

Op de webserver van een hoster wordt standaard elke benadering van de content gelogd. Als je toegang hebt tot de serverdirectory, kun je deze simpele tekstbestanden te allen tijde inzien. Bij een XAMPP-installatie schrijft Apache de gegevens bijvoorbeeld standaard in apache/logs/access.log. Een regel daarvan ziet er in z'n eenvoudigste vorm ongeveer zo uit:

```
127.0.0.1 --[10/Mar/2009:19:04:01
+0100] "GET / http/1.1" 200 5768
```

Aan het begin van de regel staat het IP-adres van de computer die het request verstuurt, in dit geval de localhost. De twee streepjes staan voor de gebruikersnamen waarmee de gebruiker op de computer is ingelogd of bij de server is aangemeld. Normaal gesproken worden hier geen gegevens ingevuld. Er volgt een tijdstempel en de eerste (en allesbepalende) regel van het HTTP-request, in dit geval een aanroep van de startpagina op de webserver. De antwoordcode '200' staat voor een correcte afhandeling. De logregel wordt tenslotte afgesloten met het aantal verstuurde bytes (zonder HTTP-overhead).

Logboek

Het Apache-configuratiebestand conf/httpd.conf slaat de gegevens van de accesslog op in log_config_module. Deze server heeft nog een ander logbestand waarin ook de bijzonder interessante inhoud van de HTTP-headers Referer en User-Agent worden opgeslagen. De logs

laten zich naar believen aanpassen en kunnen elke willekeurige HTTP-request-header opslaan.

Het merendeel van de website-exploitanten verkeert echter niet in de luxe positie dat de serverconfiguratie naar believen aangepast kan worden. Bij shared-hostingaanbieders moet je als klant genoeg nemen met wat je van de provider krijgt. Daarbij blijkt echter dat de accesslogs vaak grotendeels overeenstemmen met de voorgeconfigureerde uitvoerige Apache-variant, die als quasi-standaard geldt. Via serverapplicaties in het back-end van je website heb je dan toegang tot de logs, waarin de IP-adressen soms onherkenbaar zijn gemaakt.

Voor een statistische evaluatie heb je niet per se directe toegang tot logbestanden nodig, aangezien de meeste webhosters tools aanbieden die van de gegevensrij mooie taart- of staafdiagrammen in je browser toveren. Vaak wordt daar http-analyse voor gebruikt, een opensourceprogramma dat sinds 1996 in ontwikkeling is. De eveneens opensource Webalizer, die deel uitmaakt van XAMPP, is net zo goed als http-analyse.

Dergelijke eenvoudige analysetools laten bijvoorbeeld zien welke URL's het populairst zijn onder de bezoekers van je website. Daarbij zullen de startpagina en de bijbehorende bestanden zoals de algemene stylesheet en de favicon helemaal bovenaan staan. Interessanter is welke pagina's van daaruit het meest werden bezocht of welke rol een op de website aangeboden newsfeed speelt. Heeft iemand überhaupt die speciale link in de nieuwsbrief aangeklikt?

Hitparade

Bij het analyseren is het verschil tussen hits en page views essentieel. Een analysetool classificeert alle HTTP-requests als een hit – een enkele pagina-aanroep kan door de bijbehorende afbeeldingen dan tientallen hits genereren. De informatie die Webalizer biedt bij de categorie 'Files' is dan net wat interessanter, omdat die alleen de correct afgeleverde bestanden telt en verkeerde adressen en redirects eruit filtert.

Echt relevant zijn echter alleen de page views, die je kunt beschouwen als de verkochte op- en laagte van de website. Dat zijn de cijfers die bepalen of een pagi-

na interessant is voor reclame. Om de page views goed te bepalen, moet je meer doen dan alleen de ingebedde afbeeldingen, Flash-filmpjes, scripts en stylesheets eruit filteren. Een frameset met twee frames produceert bij de aanroep bijvoorbeeld drie hits qua aantallen HTML-pagina's, maar dat mag eigenlijk maar als één page view tellen. Webalizer kan dat onderscheidt niet maken en kent alleen 'pages', waarbij ook CSS- en scriptbestanden worden meegeteld.

Een ander relevant gegeven is het aantal visits, oftewel het aantal websitebezoeken. Om die te achterhalen, voegt de analysetool alle hits met hetzelfde IP-adres en dezelfde client binnen een bepaalde tijd bij elkaar en telt die als één visit. Daarbij kan ook de referrer gebruikt worden om afzonderlijke visits te achterhalen.

De logs kunnen niet alleen aan de hand van URL's worden bekeken, maar worden onderzocht naar de tijdstippen van de requests. Zo kun je achterhalen of de bezoekersgrafiek op bepaalde dagen duidelijk daalt. In het weekend zal de website bijvoorbeeld minder vaak worden bezocht. Uit de browserstatistieken kun je bovendien conclusies trekken over de gebruikte browsers, wat vooral handig is bij het ontwikkelen van je website: is bijvoorbeeld het testen met Internet Explorer 6 nog de moeite waard? Nadeel van http-analyse is wel dat die Internet Explorer 7 als een Mozilla-browser herkent.

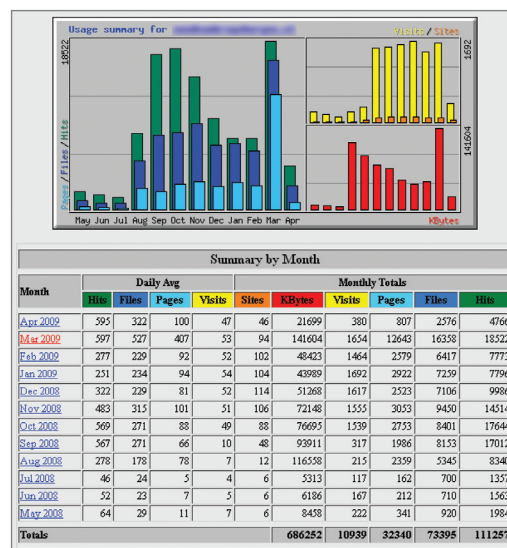
De meeste statistiektols achterhalen via het afzenderdomein waar de bezoekers vandaan komen. In het gunstigste geval

kun je daar het aandeel buitenlandse bezoekers mee bepalen, maar in het slechtste geval helemaal niets. Veel providers gebruiken namelijk een .net of .com-adres. Vandaar dat het spanner is om bijvoorbeeld de postcode van IP-adressen te achterhalen, wat bij sommige hosters mogelijk is. Dat lukt dan wel niet bij elke bezoeker, vanwege het probleem met het lokaliseren van de IP-adressen, maar levert vaak toch voldoende informatie op.

De vraag van welke website de bezoekers komen, is bijzonder interessant. Het analyseren van de referrers brengt aan het licht welke links op de website werden aangeklikt, hoeveel bezoekers door Google werden doorgestuurd en waar die bezoekers dan naar hebben gezocht. Daardoor kun je zien op welke pagina's de bezoekers binnenkomen en waar ze de site weer verlaten. Zo kom je er als eigenaar van een webwinkel bijvoorbeeld achter hoeveel geïnteresseerde kopers afhaken bij het opgeven van hun creditcardnummer en hoeveel bezoekers een link op de startpagina over het hoofd hebben gezien. Ook de gratis Webalizer, die je op access logs kunt loslaten, helpt bij zulke analyses.

Professionele trackers

Het analyseren van logfiles kan nuttig zijn voor hobby-webmasters, maar gevorderden doen liever een beroep op speciale dienstverleners. Dan moet de webmaster op elke relevante pagina een JavaScript-code invoegen, die dan weer een groter JavaScript-bestand downloadt.



Met de analysetool Webalizer krijg je een sober maar functioneel overzicht van de bezoekersgegevens, waaruit in dit geval blijkt dat maart een populaire maand was.

Time, Date Last visit Visitor type	Total access time Page impressions	IP-address Host name	Country	Origin Referrer	Browser	Browser Operating System	Screen size Colors
14:46:19 04/03/2009	1 Page	77.248.148.148 -challo.nl	Netherlands	http://www.c'ttest.nl	IE	WinXP	1280 x 1024 True Color (32 bit)
13:57:46 04/03/2009	00:00:06 4 Pages	85.148.148.148 -vanadoo.nl	Netherlands	http://www.c'ttest.nl	IE	WinXP	1280 x 1024 True Color (32 bit)
09:03:13 04/03/2009	00:00:19 3 Pages	77.248.148.148 -challo.nl	Netherlands		IE	WinXP	1280 x 1024 True Color (32 bit)
08:58:52 04/03/2009	00:00:08 2 Pages	85.148.148.148 -vanadoo.nl	Netherlands		IE	WinXP	1280 x 1024 True Color (32 bit)
16:06:40 04/03/2009	00:00:04 4 Pages	85.148.148.148 -challo.nl	Netherlands		Firefox WinXP		1280 x 1024 True Color (32 bit)
14:08:01 04/03/2009	00:12:22 12 Pages	77.248.148.148 -challo.nl	Netherlands		Opera Windows		1280 x 1024 True Color (32 bit)
14:05:55 04/03/2009	00:08:59 5 Pages	85.148.148.148 -vanadoo.nl	Netherlands		IE	WinXP	1680 x 1050 True Color (32 bit)
13:39:28 04/03/2009	00:14:03 22 Pages	77.248.148.148 -challo.nl	Netherlands		IE	WinXP	1680 x 1050 True Color (32 bit)
12:05:05 04/03/2009	1 Page	85.148.148.148 -vanadoo.nl	Netherlands		IE	WinXP	1280 x 1024 True Color (32 bit)
11:51:00 04/03/2009	00:10:45 13 Pages	85.148.148.148 -vanadoo.nl	Netherlands		IE	WinXP	1680 x 1050 True Color (32 bit)

etracker toont van iedere bezoeker alleen de helft van het IP-adres en speelt daarmee op safe in de kwestie of een IP-adres een persoonsgegeven is of niet.

Een oude bekende op dit gebied is Nedstat, die met de gelijknamige teller een van de eerste was. Met de huidige producten Sitestat en Nedstat Pro kun je het surfgedrag van websitebezoekers analyseren. Voor een prijs vanaf 170 euro per jaar krijg je dan uitgebreide rapporten om de websitestatistieken te bestuderen en daar vervolgens je site mee te verbeteren of reclamebanners op af te stemmen.

Ook etracker Web Analytics houdt alle statistieken netjes bij en wordt onder andere aangeboden als gratis Lite-versie en als Basic-versie (12 euro per maand). Met zo'n prijs is etracker Web Analytics ook betaalbaar voor kleinere websites. Daarbij zijn de versies Advanced (59 euro p/m) en Unlimited (177 euro p/m) interessant voor klanten met veel websiteverkeer.

Google Analytics is wel de bekendste webtracker. De marktleider op zoekmachinegebied distribueert Analytics sinds de overname van Urchin in 2005 gratis. Vanzelfsprekend doet Google dit niet uit pure naastenliefde. Analytics was oorspronkelijk bedoeld om het succes van AdWords-campagnes te controleren, oftewel de reclame op de Google-zoekmachine. Je kunt Analytics wel onafhankelijk van AdWords gebruiken, maar je krijgt de gratis analyse alleen in ruil voor de hiermee bij elkaar gesprokelde waardevolle gegevens. Die slaat Google op zijn eigen servers in de V.S. op – iets waar privacyactivisten en juristen niet al te blij mee zijn. Dat het gebruik van Google Analytics als illegaal zou kunnen worden gezien (zie p.90), doet overigens nauwelijks afbreuk aan diens populariteit.

Met Google Analytics (maar ook met etracker) kun je niet meer rechtstreeks bij de gegevens komen, maar worden die met een uitgebreid JavaScript (bij elke

pagina-aanroep wordt er tussen de 20 en 30 kB code in de browser geladen) uitgelezen en verstuurd. Dat heeft voor een webmaster ook voordelen: de dienst komt via het script namelijk aan informatie waar een accesslog nooit aan zou kunnen komen.

Nog gedetailleerdere informatie krijg je met heatmaps, zoals bij clickdensity.com. Dit soort diensten slaan muisklikbewegingen en/of muisklikken op en verwerken de resultaten tot een visuele weergave. Dat is met name interessant als je website nog in ontwikkeling is. Nadeel van de heatmap scripts is namelijk dat dit bij gebruikers met oudere hardware de boel aardig afremt.

Als je er niet van houdt om je bezoekersgegevens buiten de deur opgeslagen te hebben of niet rechtstreeks te kunnen benaderen – dat kan nou eenmaal bedrijfsgevoelige informatie zijn – en je de gegevens liever in een eigen database hebt staan, kun je het pakket PHP Web Stat op je eigen webserver installeren. Die moet dan wel PHP en een SQL-database ondersteunen. Ook hier moet je wat JavaScript-code aan je pagina's toevoegen, maar daarna zijn alle gegevens alleen via je eigen do-

mein en achter wachtwoord toegankelijk. PHP Web Stat is GPL en kan dus gratis gebruikt worden.

Hetzelfde geldt voor ClickHeat, waarmee je een clicks heatmap kunt maken. Je hoeft dan alleen een ZIP-bestand uit te pakken en op je server te zetten en dan de configuratiepagina config.php in je browser te openen. De versie 1.7 maakt nog geen gebruik van een (SQL-)database, dus verder hoeft je niet veel in te stellen.

Nedstat

Een van de bekendste bedrijven die zich bezighouden met websiteanalyse is het Nederlandse Nedstat. Nadat het bedrijf in juli 2005 zijn gratis Nedstat-teller aan het van origine Duitse ad pepper verkocht, biedt het nu de producten Sitestat en Nedstat Pro tegen betaling aan.

Sitestat wordt onder andere op de websites van Albert Heijn, de Volkskrant en Tweakers gebruikt. Het analyseert het surfgedrag van de websitebezoeker, het succes van onder andere banners en e-mail en meet het aantal downloads en registraties. De resultaten zijn hierbij vooral bedoeld om er marketingtechnisch beter van te worden: uit het surfgedrag van bezoekers kun je ontdekken welke pagina's het meest worden bezocht en waar reclamebanners het meest moeten opvallen. Nedstat Pro richt zich vooral op statistieken als bezoekersaantallen, waar websitebezoekers vandaan komen, welke pagina's het populairst zijn en wanneer de meeste page views zijn.

Nedstat maakt net als Google Analytics gebruik van een JavaScript-code in de HTML-pagina's, waar bij het laden van een telpixel naar verwezen wordt. In de privacy policy staat aangegeven dat de gemeten gegevens per-

soonsgegevens kunnen bevatten, zoals het IP-adres en cookies. Daar staat nadrukkelijk bij aangegeven dat die gegevens niet worden gebruikt om bezoekers te herleiden.

etracker

etracker houdt net als NedStat websitestatistieken bij door middel van een telpixel (www.etracker.de/cnt.php). Deze pixel roept het geïntegreerde trackerscript op (<http://code.etracker.com/t.js?et=...>), als er een geldige ID is meegegeven tenminste. etracker heeft in de basisversie geen cookies nodig en de transparante pixel wordt op de te onderzoeken pagina gezet. Hierbij wordt informatie ingewonnen over de beeldschermresolutie en de browsertaal. Daarbij vraagt etracker zelfs de geïnstalleerde plug-ins allemaal op.

Via het back-end kom je er achter hoeveel procent van de bezoekers ActiveX of RealPlayer kunnen gebruiken. Door een branchespecifieke referentiegroep te kiezen, kun je de resultaten vergelijken. Het systeem maakt een overzichtelijk diagram van de pagina's waarop de bezoekers binnenkomen en waarvan ze de site weer verlaten. Daarbij kun je via een enquête-tool die je op je webpagina's moet zetten de mening van gebruikers vragen. Ook kun je configureerbare rapporten laten genereren en marketingdoelen definiëren. Dat laatste blijft echter – net als de weergave per regio en steden en de analyse van mobiele internetgebruikers – voorbehouden aan de meer betalende klanten.

Met het oog op privacyrichtlijnen lijken er aan etracker wat minder haken en ogen te zitten dan aan Google Analytics, omdat de gebruikersinformatie binnen de EU blijft (het bedrijf zit in Hamburg). De detailweergave in het back-end laat slechts de helft van het IP-adres van de bezoekers zien, waardoor de gegevens je niet naar specifieke gebruikers kunnen leiden.

Google Analytics

Het Google Analytics-script (www.google-analytics.com/ga.js) maakt gebruik van een telpixel op de Google-server (www.google-analytics.com/_utm.gif) om de statistieken bij te houden. Bij het aanroepen van deze 1 x 1 pixel grote afbeelding (die alleen wordt opgevraagd, maar niet in de pagina wordt ingebouwd) verhoogt niet

Welkom op de c'ttestpagina!

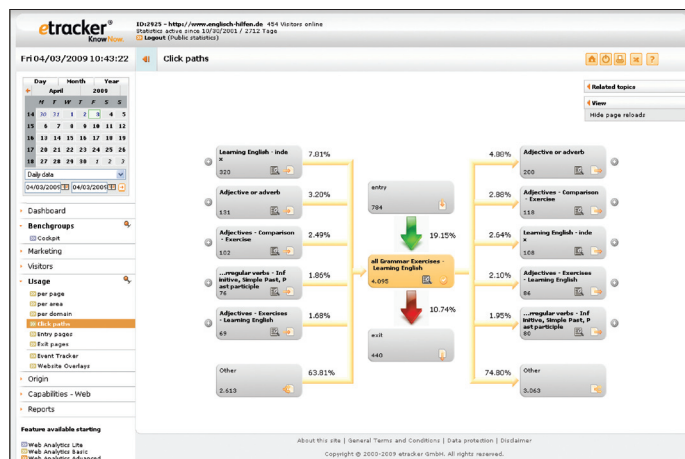
23% als je tips hebt voor het verbeteren van ons blad.

Bezoek interessante 0% over onze geteste producten.

Hier vind je allemaal 19% adgets

Bekijk hier de profielen van 13% dactie

Google Analytics geeft de interesses van bezoekers weer met behulp van een website-overlay, die de website en het klikpercentage visueel samenvoegt.



Waar komen de bezoekers vandaan? En waar gaan ze naartoe? etracker beantwoordt je vragen over het surfgedrag van website-bezoekers via een overzichtelijk diagram.

alleen de teller, maar achterhaalt ook informatie in de vorm van URL-parameters over de website en de bezoeker.

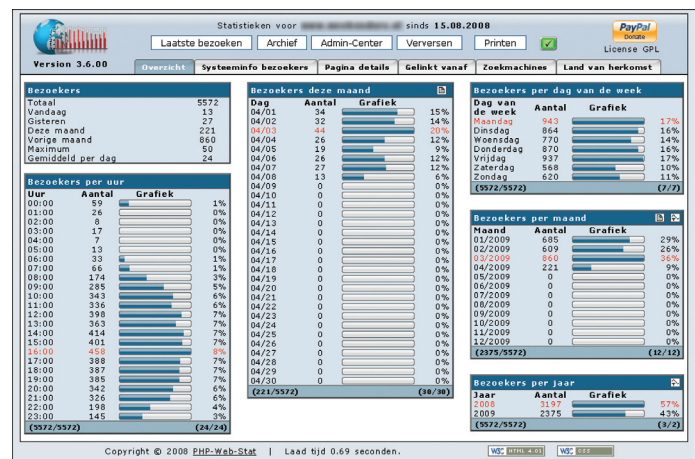
Naast de gegevens die via http verstuurd worden, weet Google dan bijvoorbeeld dat de client Nederlandstalig is, dat Analytics zich op een beeldscherm van 1680 x 1050 pixels met 32 bit bevindt en dat Java en Flash Player 10 beschikbaar zijn.

Bovendien maakt het script de cookies __utma, __utmb, __utmc en __utmz aan binnen het domein van de bezochte website. Alle vier de cookies bevatten de ID van de website-exploitant en drie ook een tijdstempel. De geldigheid van het session-cookie (__utmc) verstrijkt zodra je de browser sluit, waarna het meteen wordt verwijderd. Daarentegen verdwijnt het __utma-cookie pas na twee jaar.

Deze combinatie van cookie, telpixel en per JavaScript ingewonnen systeem-informatie geeft een opmerkelijk nauwkeurig

beeld van wat de gebruiker allemaal op de website uitvoert. Zo achterhaalt Analytics met welke zoektermen de bezoekers de site bij een zoekmachine hebben gevonden, of ze stamgasten zijn of de website voor het eerst bezoeken, hoelang ze gebleven zijn en bij welke pagina ze de site hebben verlaten. Afgezien van de technische kenmerken van de client (browser, besturingssysteem, Flash, Java, taal) achterhaalt de dienst zelfs de verbindingssnelheid. Benaderingen door zoekmachinerobots worden door het script op betrouwbare wijze uit de statistieken gefilterd.

De mooie lijn- en taartdiagrammen kunnen worden vergeleken met de gemiddelde resultaten van vergelijkbare grote websites. Analytics visualiseert de geografische afkomst van de bezoekers op een wereldkaart. De overlayweergave van de eigen website, die laat zien welke links hoe vaak werden aangeklikt, is erg fascinerend.



PHP Web Stat geeft ook uitgebreide informatie over webverkeer en bezoekers. Het pakket is opensource en kan op je eigen server geïnstalleerd worden, waardoor de gegevens niet bij derden belanden.

PHP Web Stat

De vorige oplossingen zijn allemaal van externe partijen, waardoor er extra netwerkverkeer van en naar de exploitanten van die oplossingen wordt gegenereerd. In dat verkeer zitten dan ook allerlei gegevens van de bezoekers van jouw site, en die gegevens worden dan op de servers van de dienstverleners opgeslagen. Nog geheel los van de vraag wat daar wettelijk bij is toegestaan, moet je jezelf de vraag stellen of je dat überhaupt wel wilt. Gelukkig zijn er ook mogelijkheden die informatie geheel in eigen huis te houden.

Met PHP Web Stat heb je een opensourcepakket dat makkelijk op je eigen server te installeren is. Net als bij bijna alle andere trackers voeg je een JavaScript-script toe aan je pagina's, maar dat is dan wel een aanroep binnen je eigen webserver, dus in principe niet van buitenaf in te zien. De

pagina- en bezoekersgegevens worden in een SQL-database opgeslagen en zijn dan via het backend van PHP Web Stat (in principe onder www.jouwdomein.nl/stat) met een wachtwoord toegankelijk. Er zijn algemene bezoekersgegevens, maar ook specifiek per page view, met bijbehorend IP-adres en eventuele geo-gegevens. Je kunt zien vanaf welke pagina of zoekmachine de bezoekers binnenkomen, op welke pagina ze binnenkomen en welke pagina's het vaakst bezocht zijn.

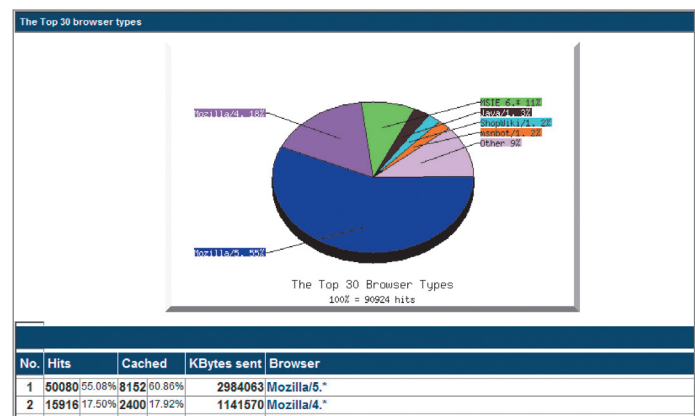
De broncode valt onder GPL, dus je kunt de source altijd nog aanpassen aan je eigen wensen als je dat wilt (en kunt).

Reclame

De hedendaagse populariteit van internet heeft een vlucht teweeggebracht in het gebruik van websiteanalyse voor reclamedoeleinden. Dat blijkt onder



Tracking voor gevorderden: heatmaps zijn zeer handig voor usability-tests.



Een marktaandeel van 75% voor Mozilla-browsers? Niet helemaal – http-analyse herkent IE7 niet correct.

andere uit een onderzoek dat onlangs in opdracht van Nedstat door Forrester Consulting in het Verenigd Koninkrijk, Frankrijk en Duitsland is uitgevoerd. Daarbij wordt gekeken naar het effect van webanalyse op e-mailmarketing. Uit dit onderzoek blijkt dat webanalyse vier keer zoveel opbrengst en achttien keer zoveel netto winst kan opleveren dan bij mailing waarbij vooraf geen webanalyse heeft plaatsgevonden [1]. Door het gedrag van potentiële klanten te analyseren, kan daar met doelgerichte mailing op worden ingesprongen en kan het succes van de marketingmethode worden gemeten. Zo kun je er bijvoorbeeld achterkomen dat een bepaalde e-mail lay-out de aandacht van de lezer veel meer trekt dan een andere of welke links het meest worden aangeklikt door welke doelgroep. Door e-mails en de lay-out van de website daar op aan te passen, kun je meer winst maken dan wanneer je geen gebruikersonderzoek uitvoert.

De marketingmensen zijn met name geïnteresseerd in gebruikersprofielen. Reclamebureaus verzamelen die dan ook

grootschalig. Google Analytics laat bijvoorbeeld zien dat webdesigners bij het tracken van hun gebruikers vaak samenwerken met hun reclameafdeling om hun website verder te verbeteren. Hoe beter een reclamemaker de bezoekers kent, des te doelgerichter de reclame kan worden aangeboden, en daarmee worden hogere inkomsten gegenereerd. In deze tijd van economische malaise is het voor adverteerders en marketeers zaak hun geld zo efficiënt mogelijk uit te geven. Daarbij wordt kennis over de mensen achter het beeldscherm steeds belangrijker.

Bedrijven als het door Google overgenomen DoubleClick, het grootste reclamenetwerk ter wereld, maken voor het inwinnen van dit soort informatie nog met name gebruik van cookies. Die registreren elk bezoek aan een pagina die op het reclamenetwerk is aangesloten, waardoor de interesses en gewoonten van de bezoeker precies in kaart kunnen worden gebracht. Daar zullen vooral de grotere bedrijven van profiteren. De rest van de internetgebruikers vraagt zich vooral af of er überhaupt nog zo-

STIR STICHTING INTERNETRECLAME		Bereik in %					
Periode: februari 2009 Bron: STIR Webmeter™		Geslacht			Leeftijd		
Domein site/pakketnaam		13+	Mannen	Vrouwen	13-34	35-49	50+
Totaal domeinen	Totaal websites	80,3	83,7	77,0	94,7	90,7	61,3
Adiab	Ajax	2,6	3,4	1,8	3,6	3,5	1,2
Adiab	Feyenoord	1,7	2,5	0,9	2,1	2,3	1,0
Adiab	PSV	1,4	2,0	0,8	1,8	2,2	0,5
Adfactor	9292ov	12,4	11,3	13,4	18,5	12,5	7,2
Adfactor	Drifzand	0,1	0,2	-	0,1	0,1	-
Adfactor	Pinin-Finifocus	4,0	4,5	3,5	4,9	5,9	1,9
Adfactor	GamingOnly	0,6	1,1	0,2	1,2	0,7	0,1
Adfactor	GasPedaal	0,9	1,4	0,5	1,1	1,4	0,4
Adfactor	Glizzy	0,2	0,3	0,1	0,2	0,3	-
Adfactor	Iens	2,6	2,2	3,0	3,2	2,8	1,9
Adfactor	ilocal	10,4	10,4	10,5	10,1	12,6	9,2
Adfactor	Juggle	0,6	0,9	0,2	0,9	0,6	0,2
Adfactor	Kaartenhuis	7,3	6,7	7,9	5,3	9,1	7,7
Adfactor	Moefz	0,1	0,1	-	0,1	0,1	-
Adfactor	MusicFromNL	1,8	1,9	1,7	2,0	2,6	1,1
Adfactor	PlanetHS	-	-	-	-	-	-
Adfactor	PocketInfo.nl	1,8	2,7	1,0	2,3	2,5	1,0
Adfactor	Retecool	0,3	0,6	0,1	0,5	0,3	0,2
Adfactor	Skoften	0,4	0,7	0,2	0,5	0,6	0,3

De STIR heeft als doel het Nederlandse internetgebruik in kaart te brengen om meer te weten te komen over de bezoekersprofielen van websites.

iets als online privacy bestaat. De meest gebruikelijke strategieën om wat privacy terug te winnen zijn het regelmatig weggooien van cookies (bijvoorbeeld bij het afsluiten van de browser) of een reclamefilter. Ook het blokkeren van cookies die niet van de website zelf afkomen (third-party cookies) is een oplossing.

Het belangrijkste voor reclamemakers blijft het aantal contacten,

in dit geval hoe vaak de reclameuitingen opgevraagd worden. Het tellen van page views en visits met behulp van serverlogs is daar niet voor geschikt, omdat dit de deur wagenwijd zou openzetten voor manipulatie. Dat heeft geleid tot het ontstaan van verschillende organisaties die de ruwe gegevens verzamelen en verder verwerken, zoals de Stichting Internetreclame (STIR).

Gebruikersstatistieken versus bescherming persoonsgegevens

Tools waarmee je webstatistieken kunt analyseren, bieden heel wat mogelijkheden om informatie uit logbestanden te halen. Daarmee kun je de toegankelijkheid en efficiëntie van je website verbeteren. Toch is niet alles wat technisch mogelijk is ook wettelijk toegestaan. Website-exploitanten bevinden zich hierdoor vaak in een grijs gebied tussen wat wel of niet mag.

In de wet staat niet waterdicht of en in welke mate webmasters informatie over websitebezoekers mogen verzamelen. De vraag is of het bij de opgeslagen en de voor de analyse verzamelde gegevens om persoonsgegevens gaat zoals in de Wet bescherming persoonsgegevens staat gespecificeerd. Volgens deze wet is een persoonsgegeven "elk gegeven betreffende een geïdentificeerde of identificeerbare natuurlijke persoon" [2]. Gegevens die niet direct naar een bepaalde persoon verwijzen, vallen dus niet onder persoonsgegevens. Denk hierbij aan statistische ge-

gevens als bezoekersaantallen van een website, informatie over websiteverkeer en dergelijke collectieve gegevens. Daarentegen zijn gegevens die vaak door de gebruiker zelf worden ingevoerd, zoals naam, geboortedatum en e-mailadres, wel persoonsgebonden.

Midden in het grijze gebied bevindt zich juist het gegeven dat het meest wordt verzameld: het IP-adres van de websitebezoeker. Er bestaat nogal wat discussie over de vraag of het hier om een persoonsgegeven gaat. Volgens het College Bescherming Persoonsgegevens (CBP) hoeft een IP-adres niet altijd als een persoonsgegeven te worden beschouwd. Wel beschouwt het CBP een volledig vast IP-adres dat door een ISP is uitgedeeld zondermeer als een persoonsgegeven, aangezien je daar een specifieke abonnee mee kunt herleiden. Hoewel je bij een dynamisch IP-adres zou kunnen denken dat het niet om een persoonsgegeven gaat, is het tegen-

deel echter waar: je krijgt van je provider wel steeds een ander IP-adres toegewezen, maar tijdelijk ben je evengoed voorzien van een specifiek adres. Volgens het CBP maakt de combinatie van een dynamisch IP-adres en het tijdstip waarop je dat gebruikte van een dynamisch IP-adres een persoonsgegeven.

Wereldwijd gezien bestaat er nog geen eenduidigheid over deze kwestie. Zo verklaarde een Duitse rechtbank vorig jaar dat een IP-adres geen persoonsgegeven is, aangezien je met een IP-adres niet zo makkelijk aan de bijbehorende abonneegegevens kunt komen. Internetproviders mogen die zonder juridische reden namelijk niet openbaar maken. Een Canadese rechter besliste in een kinderpornozaak daarentegen dat een IP-adres net als NAW-gegevens zonder rechtelijk bevel kan worden opgevraagd en te vergelijken is met een telefoonnummer in het telefoonboek. Tussen deze uiteenlopende manieren waarop met

gegevens wordt omgesprongen, bestaan natuurlijk nog allerlei uitzonderingen. Mogen persoonsgegevens wel worden vrijgegeven als je er een crimineel mee kunt opsporen of kinderporno mee kunt opdoeken? Het beslissende woord blijft hier aan de rechter. Volgens ICT-jurist Arnoud Engelfriet heeft de Nederlandse Hoge Raad bepaald dat "afgifte van NAW-gegevens door providers gevraagd mag worden als de eiser een voldoende zwaar belang heeft en er geen minder ingrijpende manier is om hem tegemoet te komen." In zaken als bovengenoemde kinderpornozaak zou je met het opsporen van de website-eigenaar kinderen en ouders zo heel wat leed kunnen besparen.

Als een IP-adres inderdaad een persoonsgegeven is, geldt dat website-eigenaars bezoekers via een privacyverklaring moeten waarschuwen wanneer een IP-adres voor gebruikersanalyse wordt gebruikt. Deze privacyverklaring moet dan direct vanaf

STIR

Bij het analyseren van het surfgedrag alleen krijg je nog geen concrete informatie over de bezoekers zelf. Maar dat is nou net de informatie waar je als bedrijf op zit te wachten om met doelgerichte reclame op in te kunnen spelen. Onder welke leeftijdscategorie vallen de meeste bezoekers, welk opleidingsniveau hebben ze en zijn het vooral mannen of vrouwen?

De Stichting Internetreclame (STIR) verricht in samenwerking met Nedstat en Intomart GfK onderzoek naar het internetgebruik in Nederland. Via de zogenaamde Webmeter wordt gekeken hoeveel mannen en vrouwen in welke leeftijdscategorie en welk opleidingsniveau welke websites bezoeken. Deze resultaten worden elke maand in een tabel op de website van de STIR gepubliceerd. Daarbij wordt alleen gekeken naar het surfgedrag op een aantal geselecteerde grote bekende websites, die specifiek deelnemen aan het onderzoek van STIR. De Webmeter is dus niet representatief voor alle websites, maar kan wel enigszins een

beeld geven van de interesses van specifieke surfercategorieën.

Er bestaat nogal wat twijfel over de betrouwbaarheid van de onderzoeksresultaten van de STIR. Zo was er onlangs enige ophef over het feit dat het bezoekersaantal van de website van de Telegraaf in oktober op onverklaarbare wijze met ruim 17% was toegenomen. Volgens de STIR lag dat aan een fout bij de redirects van twee domeinen naar pagina's van telegraaf.nl, waardoor de cijfers van de Telegraaf ten onrechte zo hoog uitvielen.

Dit voorbeeld toont duidelijk aan hoe ingewikkeld het is om websitestatistieken eenduidig bij te houden zonder dat er onjuiste gegevens insluipen. Hoe weet je of de waarden in de door de STIR samengestelde tabel representatief zijn? De resultaten komen dan wel van een panel van 8000 mensen van 13 jaar en ouder, maar dat blijft natuurlijk maar een dwarsdoorsnee van het volledige internetgebruik in Nederland. De STIR meldt dat ze in de toekomst wellicht ook het gebruik onder de 13 jaar willen gaan meten. Dat zou al een aardig

aanvulling op het totale beeld van het Nederlandse internetgebruik zijn, aangezien kinderen tegenwoordig ook flink surfen.

Privacy geen issue

De informatie die over je verzameld wordt, is indrukwekkend. Als je op je privacy gesteld bent zelfs redelijk bedreigend. Het is bijna eng om te merken dat bijvoorbeeld Amazon in diens nieuwsbrief met een verrassend aantal interessante aanbiedingen komt. Het voordeel is wel dat Amazon alleen weet hoe ons koop- en surfgedrag op hun eigen website is. Maar het wordt nog enger als reclamenetwerken in staat zijn om je over meerdere websites te volgen. Experts schatten dat DoubleClick meer dan 50 procent van alle reclamebanners via hun adserver komen. Bij iedere klik krijgen ze zo meer informatie over de gebruiker die bij het cookie hoort. En dan is DoubleClick sinds een jaar ook nog onderdeel van Google.

Er zijn dus genoeg redenen om enigszins paranoïde te worden. Het is een dunne lijn tussen het gerechtvaardigd peilen van

interesses en uithoren. De exploitanten staan daarbij onder druk omdat hun inkomsten afhankelijk zijn van de clickrates, oftewel hoe vaak er op een advertentie geklikt wordt. En die clickrates zijn de afgelopen jaren flink gedaald.

Aan de andere kant lijkt eigenlijk niemand zich zorgen te maken over de alwetendheid van bijvoorbeeld Google. Blijkbaar is het beschermen van de eigen gegevens voor de gemiddelde internetter geen thema. Er zijn een aantal wettelijke beperkingen op wat mag (zie kader), maar er blijven nog genoeg mogelijkheden om legaal informatie te verzamelen om de eigen website te verbeteren. (mda)

Literatuur

- [1] 'Integrating Web Analytics with Email Marketing to Improve Campaign Performance'. Forrester Research, Inc. <http://www.nedstat.nl/web/nedstatnl.nsf/pages/publicaties>.
- [2] Wet bescherming persoonsgegevens, Hoofdstuk 1, Artikel 1. <http://wetten.overheid.nl>.
- [3] Website Arnoud Engelfriet: <http://www.arnoud.engelfriet.net>.

de hoofdpagina van de website via een link kunnen worden ingezien en vanaf elke pagina kunnen worden gelezen. Hiermee vormen websites dus een uitzondering op de verspreiding van persoonsgegevens. Volgens de Wet bescherming persoonsgegevens mogen persoonsgegevens namelijk alleen worden verzameld wanneer de betrof fen persoon toestemming geeft tot het analyseren van deze gegevens en ook precies weet wat er met die gegevens gebeurt. De privacyverklaring moet dat dus duidelijk vermelden. Dit geldt alleen voor bedrijfswebsites en dus niet voor persoonlijke websites.

Als je bij webtracking en het analyseren van gebruikersgegevens wettelijk op safe wilt spelen, kun je bij het verzamelen van persoonsgegevens dus maar beter voorzichtig zijn. Zeker als het om IP-adressen gaat. Het anonimiseren van de gegevens is in het gunstigste geval wel toegestaan, waarbij je bij de analyse de gege-

vens niet op specifieke personen kunt betrekken.

Maar ook dan moet je de gebruiker overigens alsnog inlichten over de mogelijkheid dit verzamelen niet te accepteren (opt-out). Bovendien is het bij het analyseren verboden om anonieme gegevensverzamelingen te combineren met niet-persoonsgebonden gegevens om alsnog de identiteit van personen te achterhalen.

Aangezien een dergelijke tool momenteel niet op de markt is, bevindt het wijdverspreide gebruik van analysetools zich in een grijs gebied. Hetzelfde geldt voor hostingproviders waarbij is ingesteld dat IP-adressen volledig in de logbestanden terechtkomen. In veel gevallen levert dit gigantische gegevensverzamelingen op, waar gebruikers zonder commerciële achtergrond eigenlijk geen interesse in (zouden moeten) hebben.

Geen van de online website-analysetools is zo omstreden als

Google Analytics. Er circuleren inmiddels heel wat berichten op internet waarbij er aan getwijfeld wordt of Analytics wettelijk wel het rechte pad bewandelt. Dit komt met name doordat Analytics zowel het IP-adres als een cookie met daarin informatie over het websitegebruik van de bezoeker op een server in de VS opslaat. En dat terwijl het exporteren van gegevens naar een land buiten de EU alleen is toegestaan wanneer dit nadrukkelijk door de betrof fen persoon wordt goedgekeurd (opt-in).

Ook het feit dat Google zich het recht toeigent om bezoekersgegevens van websites te mogen gebruiken en die informatie combineert met andere gegevens, is in Europa wettelijk hoogst dubieus. Dergelijk gebruik van persoonsgegevens is alleen toegestaan wanneer elke websitebezoeker daar na een uitvoerige en begrijpelijke kennisgeving mee instemt. Daarbij is op z'n minst een informatiepagina vereist, die de bezoeker bij het benaderen

van de website voorgeschoteld krijgt en vervolgens toestemming kan geven voor het gebruik en doorgeven van zijn gegevens. Dat is momenteel niet het geval bij het gebruik van Google Analytics en daardoor is het de vraag of Analytics zich juridisch op glad ijs bevindt.

Uiteraard is het de vraag welk risico je als website-exploitant loopt als je de wetten op dit gebied aan je laars lapt. Engelfriet meldt dat je je geen zorgen hoeft te maken als je uit pure interesse statistieken van je persoonlijke website verzamelt, tenzij je die gegevens gaat publiceren. "Echte problemen kun je verwachten als je een zakelijke website hebt waar je bezoekersgedrag bijhoudt en dit doorgeeft aan de marketingafdeling of doorverkoopt aan derden. Dan is een goed privacybeleid van groot belang. Zolang je mensen vertelt wat je gaat doen, mag je veel."

Arnoud Engelfriet [3] is ICT-jurist bij adviesbureau ICTRecht.nl.

ct