

# Gedrag van consumenten

## 9. Het gebruik van clickstream data voor de analyse van consumentengedrag op het internet

G.J.L. VAN MEEER

### SAMENVATTING

Voor de meeste bedrijven en instellingen is het internet een 'black box' waarbinnen men geen grip heeft op het individuele consumentengedrag. Clickstream onderzoek is een onderzoeksmethodologie welke de 'black box' opent en inzicht geeft in het online klantgedrag. De clickstream data weerspiegelen het online klantgedrag op microniveau, per (micro)seconde wordt het klantgedrag vastgelegd. Dit artikel beschrijft het onderzoek naar de ontwikkeling en behoud van klantrelaties op een Nederlandse website, Fondsen Online (FO), waarop bezoekers onder andere informatie kunnen vinden over verschillende internationale aandelenopties en opties kopen en verkopen. Het is mogelijk de ontwikkeling en het wegloupgedrag van consumenten te volgen, als ook de bezoekers van FO op basis van hun gedrag te segmenteren. Aan de hand van een case studie demonstreren we dat bedrijven clickstream onderzoek kunnen inzetten als een effectief marketing instrument om klantrelaties te intensiveren.

**Trefwoorden: e-commerce, webhouse, clickstream onderzoek**

### 1. INLEIDING

Met nieuwe interactieve media hebben consumenten meer invloed en koopkracht (Comer et al., 1999). Echter, de meeste bedrijven en instelling weten niet met welke marketingacties zij duurzame klantrelaties kunnen opbouwen via het internet. Daarnaast zijn de meeste bedrijven en instellingen niet in staat het individuele gedrag van consumenten op het internet te analyseren. Voor hen is het internet een 'black box' waarbinnen men geen grip heeft op het gedrag van klanten en andere consumenten. Bedrijven en instellingen hebben een bruikbare onderzoeksmethodologie nodig welke gedetailleerd inzicht geeft in het klantgedrag op het internet. Dit probleem is op te splitsen in twee deelvragen:

- Wat is een bruikbare onderzoeksmethodologie om het gedrag op het internet te analyseren?
- Hoe identificeert men klantontwikkelingen en klantbehoud op een website?

- De meeste bedrijven zijn nog niet in staat om deze vragen te beantwoorden. Verschillende bedrijven gebruiken wel online rapportage tools om het bezoek op hun websites te meten. Op een relatief eenvoudige wijze kan de populariteit van de website, of delen ervan, worden bepaald. De tools zijn meestal gebruiksvriendelijk en bedrijven hoeven voor de toepassing ervan geen grote analyse-afdelingen op te zetten. Het doel van de online rapportage tools is het optimaliseren van het (verkoop-)kanaal, in dit geval de website. Er wordt onderscheid gemaakt in vier toepassingen, namelijk:
1. meten van de populariteit van (delen van) de site
  2. inzicht in de click-through ratio's in de sales funnel
  3. verbeteren van de structuur en opzet van de website
  4. opzetten van experimenten om verschillende banners of tekst frames te testen

Deze rapportage tools geven echter géén inzicht in het individuele gedrag van bezoekers op de website. Daardoor kan analyse alleen op basis van hypotheses en niet exploratief. Dit kan leiden tot het leggen van verkeerde relaties tussen variabelen en dus ook verkeerde conclusies. Daarnaast is het inzicht in het online gedrag op gegregerd niveau. Bedrijven werken veel met totalen of gemiddelden van bezoekersgedrag op de website. Doordat er geen inzicht is in het online gedrag op individueel niveau, is er dus ook geen kennis van wie klikt wanneer waarop.

De conventionele online rapportage tools kunnen dan ook niet goed worden ingezet ten behoeve van een effectieve e-commerce strategie. Weliswaar worden jaarlijks verschillende congressen en bijeenkomsten georganiseerd met sprekende termen in de titels zoals 'Web Analytics' and 'E-CRM'. Echter, weinig sprekers beschrijven de stap naar een effectieve e-commerce strategie, waarbij bezoekers via een gepersonaliseerde benadering producten of diensten worden aangeboden. Al jaren wordt Amazon.com in alle studieboeken genoemd als het voorbeeld van gepersonaliseerde verkoop, met behulp van 'collaborative filtering', via het internet. Collaborative filtering is een techniek om de interesse van een gebruiker te voorspellen aan de hand van gegevens over interesse van andere gebruikers. Helaas zijn er weinig bedrijven die zich kunnen vergelijken met de online aanbesteding- of verkooptechnieken van Amazon.com. Het internet is nog altijd een 'black box' en bedrijven kunnen hierdoor onvoldoende hun websites inzetten ten behoeve van e-commerce activiteiten. Dit artikel beschrijft op welke wijze bedrijven de stap naar gepersonaliseerde e-commerce kunnen zetten. In dit artikel wordt beargumenteerd dat dit kan aan de hand van click stream data.

In par. 2 van het artikel worden clickstream data beschreven. Aan de hand van een illustratieve case studie, in par. 3, geven we aan hoe clickstream data zijn te onderzoeken. Het artikel wordt afgesloten met een conclusie en aanbevelingen voor vervolgonderzoek (par. 4).

## 2. CLICKSTREAM ONDERZOEK

### 2.1. Clickstream data

Om de stap te kunnen maken naar gepersonaliseerde e-commerce is het noodzakelijk om clickstream data te verzamelen en analyseren. Alleen op het detailniveau

van clickstream data is het mogelijk om inzicht te krijgen in het individuele online gedrag van consumenten. Clickstream data zijn gegevens over het klikgedrag van de bezoeker op de website. Het registreert onder andere waar de bezoeker op klikt, hoe lang de bezoeker op een pagina was en in welke volgorde de bezoeker de pagina's bezocht. De clickstream data zijn dus objectief verzameld en bevatten een tijdspatroom.

Clickstream data weerspiegelen het online klantgedrag op micro-niveau; per (milj) seconde wordt het online klantgedrag vastgelegd. Het hoge detailniveau van clickstream data zorgt ervoor dat het een ongekend krachtige informatiebron is, welke gevoelig is voor elke kleine wijziging in online klantgedrag. Analyse van clickstream data voorziet onderzoekers en marketers van de mogelijkheid het bezoekersgedrag op de website te verklaren en te voorspellen. Op basis van het klikgedrag kan een analist voorspellen wat de kans is dat een bezoeker, die bepaalde pagina's bezoekt, een voor marketing gewenste handeling op de site zal uitvoeren. Hiermee is de structuur van de website te optimaliseren, om klanten in een bepaalde richting te manoeuvreren. Daarnaast kunnen bedrijven met behulp van deze data ook snel inhaken op intenties van bezoekers met interessante aanbiedingen. Op deze wijze wordt het online verkoopkanaal interactiever. De omvang en het hoge detailniveau van clickstream data lijken een nieuw vakgebied te creëren binnen (markt-)onderzoek. Echter, de huidige statistische technieken zijn ook toepasbaar op deze data. Met exploratief onderzoek en 'data mining' is het mogelijk om zonder hypothesen op zoek te gaan naar verborgen relaties.

Een voordeel van clickstream data is dat het inzicht geeft in online gedrag op individueel niveau van bezoekers op de website. De analist weet precies wie, wanneer en waarop klikt. Hierdoor kunnen tijdens de analyses allerlei mogelijke dwarsdoorsneden van de bezoekers worden gemaakt en kunnen verschillende groepen gebruikers worden gesegmenteerd. Er kunnen tevens verschillende gedragsmodellen worden ontwikkeld, zoals verkoop-, retentie- en/of churn modellen. Indien bezoekers in 'secure sites' herkend kunnen worden, is het een mogelijkheid om clickstream data te verbinden met klantgegevens uit de marketing database. Met deze datafusie hebben bedrijven een integraal klantbeeld van het offline en het online gedrag. Kortom, clickstream onderzoek kan fungeren als een instrument om het klantgedrag te sturen.

Het opzetten van clickstream onderzoek kent een paar uitdagingen. Zo is het verzamelen van clickstream data en het opbouwen van een webhouse, een datawarehouse (DWH) met clicks, een relatief nieuw vakgebied. Een website wordt beheerd op een 'server'. Een bezoeker van een website zoekt verbinding met deze server. De server verstuurt kleine bestandjes (html, gif) naar de browser van de bezoeker, en zodoende wordt een webpagina opgebouwd en getoond. De server logt informatie, ook wel clickstream data, over elke interactie tussen de server en de gebruiker. Met deze server data is het mogelijk om een webhouse op te bouwen. Daarnaast is het mogelijk om een webhouse op te bouwen met clickstream data in samenwerking met een Application Server Provider (ASP); dit zijn dezelfde bedrijven die de online rapportages tools aanbieden. Een ASP plaatst voor haar online rapportage tools onzichtbare tags op elke pagina, tekstframe en banner in de website. Deze tags loggen elke interactie tussen de bezoeker en de website. Hiermee is het mogelijk om een logfile met clickstream data te creëren.

Clickstream data kennen verschillende componenten (Martison, 1999; Mena, 1999;

Kimball & Merz, 2000; Linoff & Berry, 2002). Welke componenten worden gelogd kan bepaald worden met de configuratie van de server. Belangrijke componenten van clickstream data zijn het IP-adres, cookie, datum en tijd, de betreffende pagina (Transaction log), vorige pagina (Referral log), en browser type (Agent log). Tabel 1 geeft een voorbeeld van een bestand met clickstream data.

Tabel 1. Voorbeeld van een bestand met clickstream data.

ip address_cookie	date_time	method	referrer	status	ip	browser
62.166.17.62.29050106764529036	[0]/Nov/2003/00:08:10	GET	/clickstream?zoom&bandid=1480&sort	200	http://www.internetontgereguleer.nl/14	Windows2000
62.166.17.62.29050106764529036	[0]/Nov/2003/00:10:01	GET	/14/default.html?	200	http://www.internetontgereguleer.nl/14	Windows2000
62.166.17.62.29050106764529036	[0]/Nov/2003/00:11:25	GET	/doelhoeveelwerk.html?	200	http://www.internetontgereguleer.nl/14	Windows2000
80.61.105.112.1444910644740642	[0]/Nov/2003/00:23:10	GET	/clickstream?zoom&bandid=1480&sort	200	http://www.internetontgereguleer.nl/14	Windows2000
80.61.105.112.1444910644740642	[0]/Nov/2003/00:25:50	GET	Anonymous	200	http://www.internetontgereguleer.nl/14	Windows2000
81.59.9.97.280701067185542568	[0]/Nov/2003/00:37:47	GET	/clickstream?zoom&bandid=1480&sort	200	http://www.internetontgereguleer.nl/14	Windows2000
62.59.144.37.57291067647027141	[0]/Nov/2003/00:38:09	GET	/14/default.html?	200	http://www.internetontgereguleer.nl/14	Windows2000
80.61.105.112.1444910644740642	[0]/Nov/2003/00:38:17	GET	/clickstream?zoom&bandid=1480&sort	200	http://www.internetontgereguleer.nl/14	Windows2000
80.61.105.112.1444910644740642	[0]/Nov/2003/00:38:41	GET	/16/default.html?	200	http://www.internetontgereguleer.nl/16	Windows2000
62.59.144.37.57291067647027141	[0]/Nov/2003/00:38:42	GET	/doelhoeveelwerk.html?	200	http://www.internetontgereguleer.nl/14	Windows2000
80.61.105.112.1444910644740642	[0]/Nov/2003/00:38:52	GET	/doelvoorbeeld.html?	200	http://www.internetontgereguleer.nl/14	Windows2000
213.17.23.97.28110676471212199	[0]/Nov/2003/00:39:19	GET	/clickstream?zoom&bandid=1480&sort	200	http://www.internetontgereguleer.nl/16	Windows2000

## 2.2. Valkuilen

De kwaliteit van clickstream data moet uiteraard hoog zijn. Hiervoor is een zeer nauwkeurig ETL (Extraction, Transformation and Loading) proces vereist. Data extractie betreft het ophalen van de data bij de bron. Op de data vinden enkele transformaties plaats, zoals codering van variabelen, zodat het tot een bruikbaar databestand wordt geconverteerd. Vervolgens moeten de data in een bepaalde structuur in de webhouse worden geladen (Sweiger et al., 2002).

Een belangrijk onderdeel van het ETL proces is het omgaan met fouten of incomplete data (Bauista & Silva, 2001). Het eerste probleem is caching. Bezochte pagina's worden opgeslagen in de cache om de uitwisseling tussen de browser en de server te reduceren. Als de gebruiker de "back" -button aanklikt dan wordt de opgeslagen pagina getoond, en de server registreert niet het herhalingsbezoek van de pagina (Mobasher et al., 2000). Een ander probleem wordt veroorzaakt door 'web spiders', die de website doorzoeken. Dit zijn robots, die niet-menselijke sporen achterlaten en misleidend zijn bij het analyseren van surfgedrag (Spiliopoulou, 2000). Daarnaast geven clickstream data afkomstig van proxy servers problemen bij het identificeren van website gebruik (Cooley, 1999). Een proxy server regelt het verkeer tussen de server van de website en de browsers van een groep gebruikers. Hierdoor worden werknemers van (grote) bedrijven in de logfiles geregistreerd als één gebruiker.

Zonder een unieke gebruikersnaam of zonder het toekennen van cookies is het erg moeilijk om unieke gebruikers te identificeren (Mena, 1999). Een cookie is een code, die ongemerkt gekoppeld wordt aan de browser van de surfende consument. Elke keer dat de bezoeker de website bezoekt, herkent de server de browser van de gebruiker via de cookie. Steeds meer websites maken een onderscheid tussen een open en een gesloten omgeving bij het aanbieden van informatie en diensten. Indien de website een 'gesloten' omgeving heeft en bezoekers dienen in te loggen

met een wachtwoord (bijvoorbeeld internet bankieren), dan kunnen we naar persoonlijke informatie vragen en deze vasthouden. De persoonlijke informatie geeft voor analyse doeleinden meer mogelijkheden. Naast persoonlijke informatie kunnen ook het huidige productiebezit, bijvoorbeeld de klant heeft twee spaarproducten en drie beleggingsfondsen, of andere gegevens koppelen aan zijn/haar klikgedrag. Als we weten welke klant bij welk surfgedrag hoort dan is dit gedrag te koppelen aan andere gegevens in de marketingdatabase.

De hoeveelheid en het hoge detailniveau van clickstream data maken het moeilijk voor datawarehouse managers van bedrijven om een webhouse op te bouwen. Vaak zijn zij onbekend met dit type data. Veel managers willen dergelijke grote hoeveelheden clickstream data ook niet onderhouden. Voor veel onderzoek moeten de data enkele maanden worden bewaard en voor bepaalde longitudinale toepassingen zelfs langer. ASP's hebben de benodigde expertise en in samenwerking is het mogelijk om een webhouse met clickstream data op te bouwen.

Een ander probleem is dat het gebruik van clickstream data voor verkoopdoeleinden schuurt tegen de privacygevoeligheid. De meeste consumenten zijn zich niet bewust dat hun surfgedrag wordt vastgelegd en geanalyseerd. Veel banken hebben, bijvoorbeeld, een 'secure' internet bankieren site, waarop hun klanten inloggen en transacties kunnen doen. Deze klanten zijn herkenbaar via hun gebruikersnaam en cookies. Als zij zogenaamd anoniem de 'public' site bezoeken, zijn zij wel herkenbaar via hun cookie. Echter, de bank mag dit niet gebruiken om deze klanten een product aanbod te doen op basis van hun surfgedrag op de 'public' site. Een disclaimer op de website kan in bepaalde mate tegemoet komen aan de privacy van het klikgedrag van bezoekers.

## 3. CLICKSTREAM ONDERZOEK IN DE PRAKTIJK

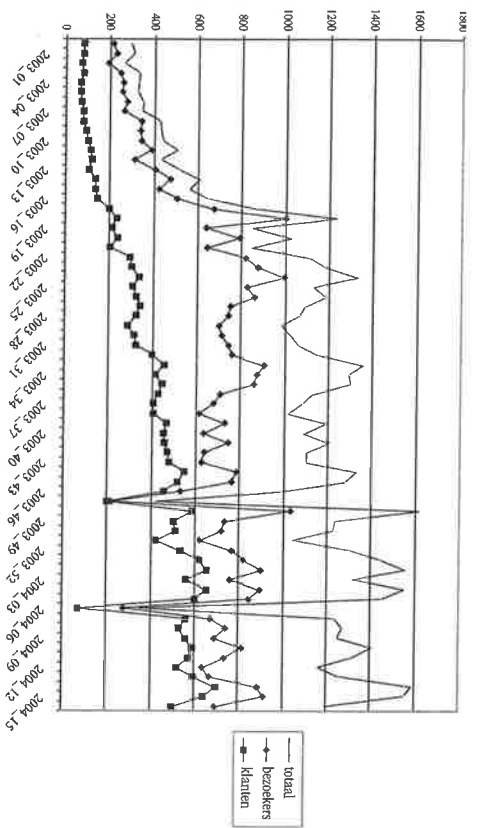
### 3.1. Case study: 'Fondsen Online'

Met de direct marketing concepten, zoals 'het creëren en behouden van klantrelaties', hebben marketeers een beter begrip van klantrelaties op een website. Hieronder wordt een voorbeeld van clickstream onderzoek naar de ontwikkeling en behoud van klantgedrag op een website van een financiële instelling in Nederland beschreven. Via deze website kunnen bezoekers zich aanmelden als klanten en in een persoonlijke omgeving internationale opties vergelijken en verhandelen.

'Fondsen Online (FO)' is een website waar klanten een rekening kunnen openen om fondsen aan te kopen en te verhandelen. Zowel consumenten als professionele brokers gebruiken FO vanwege de goede informatievoorziening en de relatief eenvoudige en goedkope wijze van fondsen aan- en verkopen. Bezoekers van FO hebben de keuze om een rekening te openen en fondsen te verhandelen of niet. FO kent twee type gebruikers: bezoekers en geregistreerde klanten. Het gedrag van een bezoeker wordt gelogd via een cookie en blijft anoniem. Een geregistreerde klant uiteraard te volgen via een cookie en gebruikersnaam.

Voor onderzoeksdoeleinden is clickstream data van januari 2003 tot en met april 2004 verzameld. Gedurende deze periode zijn er ruim 3,6 miljoen pagina's van FO bezocht in ongeveer 460.000 sessies. Marketeers wilden weten hoe klanten zich in

loop van de tijd ontwikkelen en welke klanten wegglopen. De activiteit van klanten op FO is per week bekeken. Figuur 1 laat het aantal bezoekers per week zien in de periode januari 2003 tot en met april 2004. Er is een duidelijke toename in zowel bezoekers als klanten te zien.

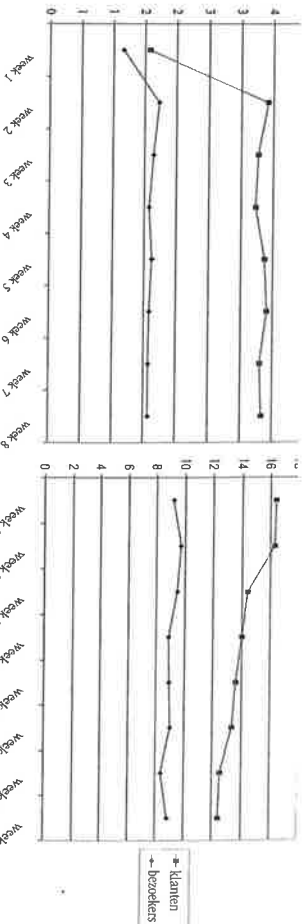


Figuur 1. Aantal bezoekers per week op FO in de periode januari 2003 tot en met april 2004.

### 3.2. Klantonwikkeling

Om klantrelaties op te bouwen moet de analist achterhalen welke gebruikers zich ontwikkelen tot actieve klanten. Met clickstream onderzoek kan er longitudinale analyse naar de ontwikkeling van online gedrag worden uitgevoerd.

Hieronder wordt het verschil in het gemiddelde aantal sessies in de eerste acht weken weergegeven tussen *bezoekers* die bezoekers blijven na acht weken en bezoekers die *klant* worden (zie figuur 2.1). Een 'klant' is een bezoeker die zich heeft geregistreerd en minimaal één order heeft uitgevoerd. In figuur 2.2 wordt het verschil in het gemiddelde aantal bezochte pagina's per sessie tussen bezoekers en klanten van FO in de eerste acht weken weergegeven. Diegenen die klant worden hebben gemiddeld een hoger aantal sessies per week en bekijken gemiddeld meer pagina's per sessie dan bezoekers die géén klant van FO worden. Dit is al meteen in de eerste twee weken duidelijk zichtbaar. Het verschil in het aantal sessies ( $F=10.66, p<0.001$ ) en het aantal bezochte pagina's ( $F=3938.63, p<0.001$ ) tussen bezoekers en klanten is significant. Er is ook een significant interactie-effect tussen type gebruiker (bezoeker of klant) en tijd (weken) voor het aantal sessies ( $F=2.82, p=0.006$ ) en voor het aantal bezochte pagina's per sessie ( $F=274.04, p<0.001$ ). Het laatste betekent dat het type gebruiker verschilt in gedrag in de loop van de tijd. Het belangrijkste is dat de figuren indicaties geven voor het herkennen van potentiële klanten onder de bezoekers. Vooral de toename in sessie onder klanten in de eerste twee weken (figuur 2.1) is hierbij relevant. Dit soort bezoekers kunnen op een andere manier worden benaderd, via banners e.d., dan de bezoekers die deze toename niet vertonen.



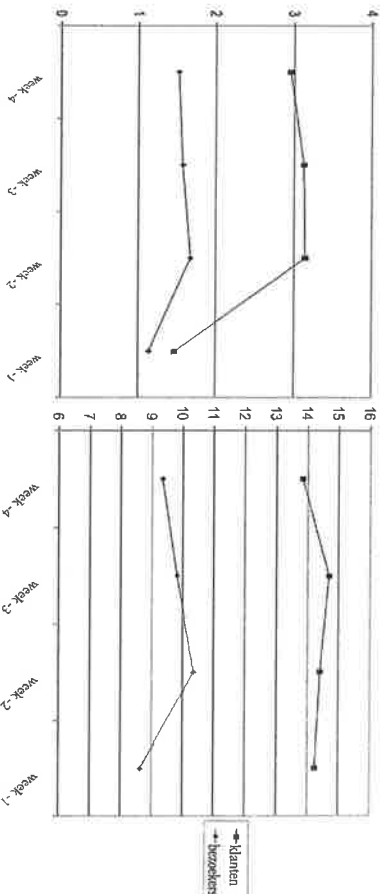
Figuur 2.1. (links) Aantal sessies van bezoekers en klanten in de eerste 8 weken.

Figuur 2.2. (rechts) Aantal pagina's per sessie van bezoekers en klanten in de eerste 8 weken.

### 3.3. Klantbehoud

Niet alleen het aantrekken van nieuwe klanten is van belang, ook het behoud van klanten dient te worden onderzocht. Een belangrijke vraag hierin is de volgende: Is het mogelijk om potentiële wegglopers te identificeren op basis van hun gedrag in de laatste weken dat zij de site bezoeken?

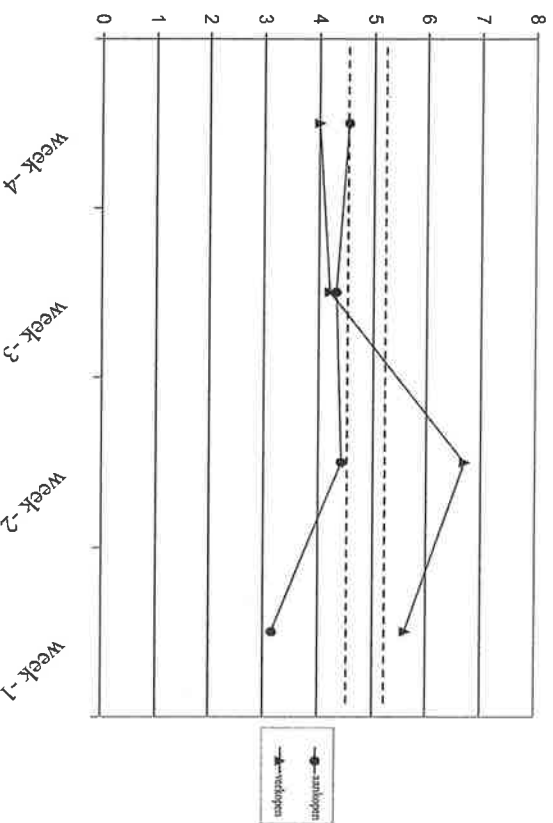
Met longitudinale analyses op clickstream data is het mogelijk om inactieve gebruikers te herkennen. Hiervoor kijken we weer naar het gemiddelde aantal sessies en het aantal opgevraagde pagina's per sessie van bezoekers en van klanten in de laatste vier weken op FO. Week -1 betekent de laatste week op FO. In figuur 3.1 is het gemiddelde aantal sessies per week van zowel bezoekers als klanten weergegeven. In figuur 3.2 is het gemiddelde aantal pagina's per sessie van zowel bezoekers als klanten weergegeven. Het verschil in het aantal sessies en het aantal pagina's per sessie tussen beide gebruikersgroepen is wederom onderzocht.



Figuur 3.1. (links) Gemiddeld aantal sessies van bezoekers en klanten gedurende de laatste 4 weken.

Figuur 3.2. (rechts) Gemiddeld aantal pagina's per sessie van bezoekers en klanten gedurende de laatste 4 weken.

Het blijkt dat er een significant verschil is in het aantal sessies ( $F=42.94$ ,  $p<0.001$ ) en het aantal opgevraagde pagina's per sessie ( $F=32.078$ ,  $p<0.001$ ) tussen beide gebruikersgroepen. Figuur 3.1 laat een duidelijke terugval zien bij beide gebruikersgroepen in de laatste week. Vooral voor klanten is een terugval in het aantal sessies in de laatste week goed waar te nemen; wie hoog zit, kan diep vallen. Er is echter géén significant interactie-effect waar te nemen tussen type gebruiker (bezoeker of klant) en tijd (weken) voor zowel het aantal sessies als voor het aantal pagina's per sessie gedurende de laatste weken op FO. Dit betekent dat op basis van activiteit (hoe vaak men klikt) het moeilijk is om het wegloupedrag te voorspellen. Misschien is dit anders als we kijken naar welke type bezochte pagina's (waarop men klikt) gedurende de laatste 4 weken. Hiervoor kijken we in figuur 4 naar het aantal keer dat de klant (op de bevestigingspagina heeft geklikt nadat hij of zij) een beleggingsfonds heeft gekocht dan wel heeft verkocht in de laatste vier weken. De gestippelde lijnen zijn de gemiddelden.



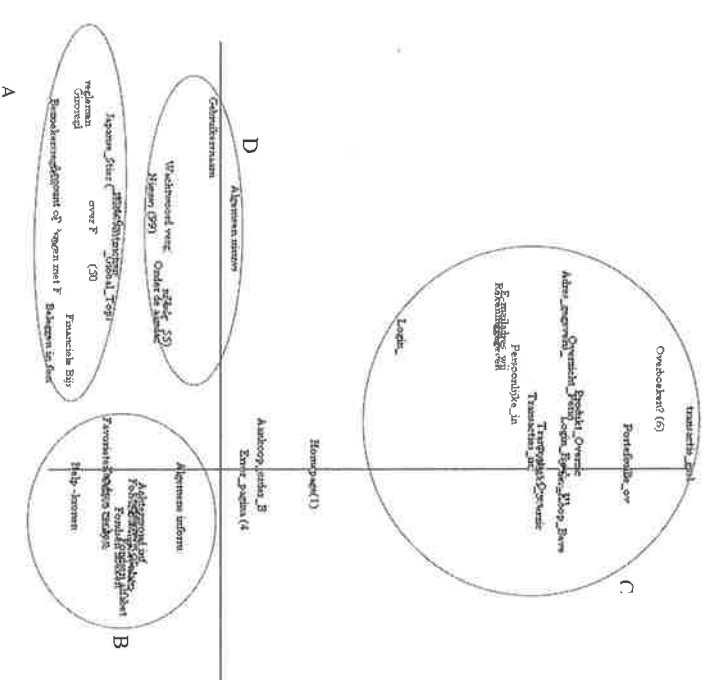
Figuur 4. Gemiddeld aantal keren dat klanten beleggingsfondsen kopen en verkopen in de laatste vier weken.

Figuur 4 laat zien dat het aantal aankopen in de laatste weken afneemt. Het aantal verkopen van beleggingsfondsen stijgt in de laatste twee weken dat men nog een actieve klant is bij FO. Weglopers verkopen alle beleggingsfondsen voordat ze inactief worden. FO zou met een webhouse een 'early warning system' kunnen opzetten, waarbij weglopers in een vroeg stadium aan hun bezoekersgedrag herkend kunnen worden. Klanten die qua activiteit onder het gemiddelde zakken, kunnen door FO individueel benaderd worden met een gepaste marketing actie, bijvoorbeeld een aanbieding met korting op het aankopen van de volgende vijf beleggingsfondsen. vervolgonderzoek zal moeten bewijzen of 'off line' kenmerken, zoals waarde van het fonds, expiratie datum en/of dividend een relevante rol speelt in klantontwikkeling en -behoud op een website.

### 3.4. Segmentatie

Niet alleen het aantrekken en behouden van klanten kan worden ondersteund met click stream data. De volgende vraag kan ook worden beantwoord met behulp van deze data: Welke verschillende behoeften hebben bezoekers van FO? Voor een antwoord op deze vraag, moet gekeken worden naar combinaties van pagina's die worden bezocht binnen dezelfde sessie. Als er duidelijke groepen van pagina's zijn, dan is het mogelijk om segmenten van bezoekersgedrag te bepalen. Als we veronderstellen dat men met een bepaalde intentie een website bezoekt, dan is elk segment gelijk aan de behoefte van een groep gebruikers.

Aangezien een website uit veel pagina's bestaat, wordt eerst een Correspondentie analyse uitgevoerd. Met een Correspondentie analyse op de matrix die de aantallen bevat van het voortkomen van combinaties van pagina's binnen een sessie is het mogelijk om unieke typologieën van sessies te ontdekken. Correspondentie analyse visualiseert een verborgen structuur in de datamatrix (Benzécri, 1973; Greenacre, 1984). Een gegeven inertie stelt vast of een bepaalde weergave acceptabel is, en wat de waarde van de assen (proportionele variantie uitgedrukt in TAU) zijn voor de totale oplossing (Hoffman & Franke, 1986). In figuur 3.4.1 is een oplossing gegeven (dimensie 1,2; cumulatieve Tau = 0,61). Hoe dichter de pagina's 'bij elkaar liggen', des te vaker worden ze in combinatie binnen een sessie bezocht.



Figuur 5. Correspondentie analyse (CORWIN) van FO (dimensie 1, 2)

Uit figuur 5 blijkt dat er vier groepen pagina's binnen dezelfde sessie vaak in combinatie worden opgevraagd: Type A: pagina's om informatie over de FO website te



verkrijgen (informatie site); Type B: pagina's om actuele informatie over fondsen te verkrijgen (informatie over producten); Type C: pagina's waarin fondsen worden verhandeld (uitvoeren transacties); Type D: pagina's om informatie over actuele onderwerpen (nieuws). Een veel voorkomende segmentatie bij websites is een segmentatie gebaseerd op het verschil tussen klanten die de website primair gebruiken voor het uitvoeren van transacties, versus diegenen die de site gebruiken voor informatiewinning (Kalakota & Whinston, 1996). Bovenstaande indeling is hier consistent mee. Net als in andere toepassingen van segmentatieonderzoek is een verdere profilering mogelijk. In tabel 2 is de gemiddelde duur en het gemiddelde aantal pagina's per sessie van de vier typen weergegeven.

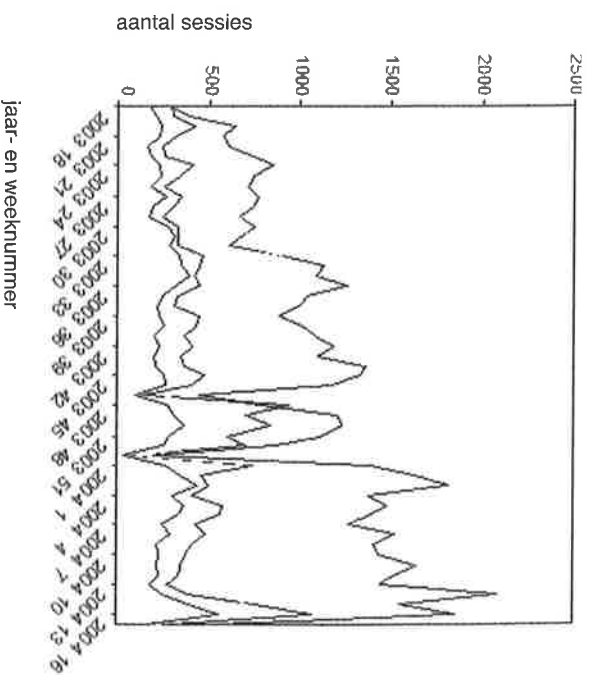
Tabel 2. De gemiddelde duur en het gemiddelde aantal pagina's per sessie van type A, type B, type C en type D.

	type A	type B	type C	type D
gemiddelde duur (minuten)	11,8	11,2	10,2	11,7
gemiddeld aantal pagina's	17,8	17,1	13,8	17,1

Het blijkt dat sessies waarbij informatie wordt opgevraagd (type A, B en D), sessies zijn waarbij bezoekers relatief ook meer pagina's aanklikken. Tijdens sessies, waarbij een order wordt uitgevoerd (type C), worden relatief minder pagina's aangeklikt en deze sessies duren ook minder lang.

Nadat met Correspondentie analyse groepen van pagina's zijn gevonden, is het mogelijk om een Cluster analyse uit te voeren op de gevonden groepen pagina's. Hieruit blijkt welke combinaties van groepen pagina's mogelijk zijn. Figuur 6 geeft het verloop van de drie populairste sessies in de loop van de tijd weer: 34,7% van de sessies met louter type C (uitvoeren transacties), 10,2% van de sessies met type B, C (informatie over producten, uitvoeren transacties) en 7,2% van louter sessies met type B (informatie over producten).

Figuur 6 laat zien dat sessies waarin louter effecten orders worden uitgevoerd (type C) het meeste plaatsvinden. Het aantal sessies met type C is gedurende de onderzoeksperiode ook aanzienlijk gestegen. Daarnaast blijkt dat er sessies zijn waarin bezoekers zowel informatie over de beleggingsfondsen bekijken (type B) als ook effecten orders uitvoeren (type C). FO moet echter de bezoekers die louter informatie opzoeken over beleggingsfondsen (type B) benaderen en overhalen, met bijvoorbeeld een introductiekorting, om ook via FO transacties uit te voeren. Voor typen B, C en C zou een bedrijf vooral een retentiebeleid kunnen ontwikkelen. Andere consequenties voor marketing zijn ook denkbaar, maar vallen buiten het bestek van dit artikel.



Figuur 6. Verloop van het type sessies in de periode april 2003 t/m april 2004.

#### 4. CONCLUSIES EN AANBEVELINGEN

Voor winstgevendheid op de lange termijn is klantbehoud van groot belang (Wiegman & Koth, 1999). Als bedrijven retentie willen verhogen, dan moeten zij klanten 'opsluiten' in een proces dat moeilijk is te verlaten (Nemzow, 1999). Een website moet een lange termijn 'stickiness' bewerkstelligen en een financiële drempel creëren dat klanten ontmoedigt om over te stappen naar de concurrentie (Nemzow, 1999). Veel websites zijn succesvol in het aantrekken van veel bezoekers, maar zij hebben moeite om hen om te vormen tot bezoekers of klanten met herhaal aankopen (Agrawal et al., 2001). Het bouwen van een winstgevende website met een vaste groep klanten is ingewikkelder dan het ontwerpen van een paar mooie webpagina's en maar wachten tot het geld binnen rolt (Wiegman & Koth, 1999). Veel bedrijven slagen er niet in om de website te integreren in de marketing strategie, omdat zij geen duidelijke strategie hebben op het gebied van klantontwikkeling en -behoud (Van Meer, 2006a). Er zijn verschillende mogelijkheden om het klantgedrag te sturen en de winstgevendheid van de klant te optimaliseren. De interactieve mogelijkheden van het internet helpt bij het creëren en het verbeteren van de relaties met klanten. Alles wat digitaal is kan gepersonaliseerd worden, en op termijn zullen ook grote instellingen relaties ontwikkelen met individuen (Pine et al., 1995). Echter, verscheidene interacties en structureel contact via het internet zijn nodig om een relatie te creëren en te behouden tussen bedrijven en hun klanten. Het behouden van klantrelaties moet ook deel zijn van de marketing strategie.

Uit de case studie blijkt dat clickstream onderzoek geschikt is om de ontwikkeling van online klantgedrag in de tijd te analyseren. Door identificatie van potentiële loyale klanten draagt clickstream onderzoek bij aan inzicht in de ontwikkeling

van klantrelaties. Tevens kan elke negatieve verandering in het gedrag van potentiële weglopers vroegtijdig worden gesignaleerd. Met een webhouse kunnen alle klanten met een onder gemiddelde activiteit direct worden benaderd om dit negatieve gedrag te onderbreken en wenselijk gedrag te stimuleren.

Als de technologische ontwikkeling zich voortzet, wat waarschijnlijk is, dan wordt de behoefte aan nieuwe mogelijkheden om het online gedrag te analyseren en te modelleren groter. Online rapportage tools kunnen een meerwaarde bieden bij het meten van website bezoek en het optimaliseren van het online verkoopkanaal. Echter, om de stap te maken naar gepersonaliseerde (one-to-one) marketing is een operationele webhouse noodzakelijk. Klanten hebben al jaren hooggespannen verwachtingen, maar vele bedrijven hebben moeite om deze stap te zetten. Bedrijven en instellingen hebben de uitdaging om de verwachtingen waar te maken en op-maat dienstverlening via het internet aan te bieden. Clickstream onderzoek maakt gepersonaliseerde benadering mogelijk door het online gedrag op individueel niveau nauwkeurig te analyseren en de aanbeveling- en verkoopmodules op de website te voeden met online gedragsmodellen.

De case studie die in dit artikel is gerapporteerd toont aan dat analyse van clickstream data het mogelijk maakt om een scherp inzicht te krijgen in het individuele gedrag van bezoekers op een website. Analyse van deze data kan worden gebruikt om te voorspellen wie er klant wordt van een site of juist weggaat en voor het segmenteren van klanten. Er zijn ook andere analysetechnieken bruikbaar, zoals (Van Meer, 2006b):

- Relatie tussen het online bezoekersgedrag en de KPI's met regressieanalyse;
  - Segmentatie van sessies van bezoekers met cluster analyse.
- Vervolgonderzoek kan zich verder richten op het benutten van bestaande onderzoekstechnieken gebruikmakend van clickstream data, bijvoorbeeld:
- Niet elke sessie verloopt succesvol met bijvoorbeeld een aankoop als eindresultaat. Om de verschillende klikpatronen van succesvolle sessies te onderscheiden van niet-succesvolle sessies kan een classificatie techniek, zoals CHAID (Chi-square Automatic Interaction Detector), worden toegepast voor dit doelinde.
  - Een sessie is veelal onderdeel van een reeks van sessies. Om inzicht te krijgen hoe gedrag zich in de loop van de tijd ontwikkelt, sequentiële patronen van sessies kunnen onderzocht worden met Markov Modellen.

De onderzoeksresultaten en de uitkomsten van de voorspellingsmodellen zijn geschikt voor gepersonaliseerde aanbevelingen voor informatie, producten of diensten. De komst van een nieuwe data bron vraagt niet altijd acuut om een nieuw onderzoeksprogramma (Bucklin et al., 2002), maar het toenemende gebruik van websites en de daarmee gepaarde groei in de hoeveelheid beschikbare informatie stellen nog een aantal belangrijke uitdagingen voor onderzoekers die clickstream data ook op de juiste waarde weten te schatten.

## LITERATUUR

- Agrawal, V., L.D. Arjona, & R. Lemmens (2001). E-performance: The Path to Rational Exuberance. *The McKinsey Quarterly*, No 1, 31-43. Retrieved from: <http://www.mckinseyquarterly.com>.
- Baista, P. & M.J. Silva. (2001). Mining Web Access Logs of an Online Newspaper. In: *Proceedings of 12th International Meeting of the Euro Working Group on Decision Support Systems*, Cascais (Portugal).
- Benécir, J.-P. (1973). *L'analyse des données II. [Data Analysis II] L'analyse des correspondances [Analysis of Correspondences]* - third edition. Paris: Dunod.
- Bucklin, R.E. & C. Simeiro, (2002). A Model of Website Browsing Behaviour. *Journal of Marketing Research*, (August), 249-267.
- Comer, J.M., R. Mehta, & T.L. Holmes (1999). Information technology: Retail users versus nonusers. *Journal of Interactive Marketing*, 12 (2), 49-62.
- Coolley, R., B. Mobasher & J. Srivastava (1999). Data Preparation for Mining World Wide Web Browsing Patterns. *Journal of Knowledge and Information Systems*, 1 (1), 11-15
- Greenacre, M.J. (1984). *Theory and Applications of Correspondence Analysis*. London: Academic Press.
- Hoffman, D. & G. Franke (1986). Correspondence Analysis: Graphical Representation of Categorical Data in Market Research. *Journal of Marketing Research*, 23, 213-227.
- Kalakota, R. & A.B. Whinston (1996). *Frontiers of Electronic Commerce* (1st ed). Reading (Mass) Addison-Wesley Pub.
- Kimball, R., & R. Merz (2000). *The Data Warehouse Toolkit: Building the Web-Enabled Data Warehouse*. New York: John Wiley.
- Linoff, G.S. & M.J.A. Berry (2002). *Mining the Web: Transforming Customer Data into Customer Value*. New York: John Wiley.
- Mattison, R. (1999). *Web Warehousing and Knowledge Management*. New York: McGraw-Hill.
- Meert, G.J.L. van & W.F. van Raaij (2004). A Suitable Research Methodology for Analyzing Online Banking Behaviour. *Journal of Internet Banking and Commerce*, 9, (1). Retrieved from: <http://www.arraydev.com/commerce/jibc/0402-04.htm>.
- Meert, G.J.L. van (2006a). Customer Development and Retention on a Web-banking Site. *Journal of Interactive Marketing*, 20 (1) 58-64.
- Meert, G.J.L. van (2006b). *Financial Behaviour on the Internet*. Alblassterdam: Haveka BV.
- Mena, J. (1999). *Data Mining your Website*. Boston: Digital Press.
- Mobasher, B., R. Coolley & J. Srivastava (2000). Automatic Personalization Based on Web Usage Mining. *Communications of the Association for Computing Machinery*, 43 (8) 142-151.
- Nemzow, M. (1999). E-commerce "Stickiness" for Customer Retention. *Journal of Internet Banking and Commerce*, 4, no. 1. Retrieved from: <http://www.arraydev.com/commerce/jibc/9908-03.htm>
- Pine, B.J. II (1993). *Mass Customization - The New Frontier in Business Competition*. Boston: Harvard Business School Press.
- Spiliopoulou, M. (2000). Web Usage Mining for Websites Evaluation. *Communications of the Association for Computing Machinery*, 43 (8), 127-134.
- Sweiger, M., M.R. Madsen, J. Langston & H. Lombard (2002). *Clickstream Data Warehousing*. New York: John Wiley.
- Wiegman, G. & H. Koth (1999). Customer Retention in Online Retail. *Journal of Internet Banking and Commerce*, 4 (1). Retrieved from: <http://www.arraydev.com/commerce/jibc/9909-07.htm>