

HET VERHOGEN VAN PERFORMANCE-TOLERANTIE IN WEBSHOPS

STAGEVERSLAG MASTER BUSINESS ANALYTICS

Auteur

Lidewij Kooiman

Begeleiders

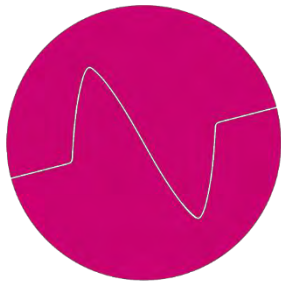
ir. Raoul Wilmans (MeasureWorks)

dr. Sandjai Bhulai (VU)

dr. Eduard Belitser (VU – tweede lezer)

MeasureWorks, Almere

Vrije Universiteit, Amsterdam



MeasureWorks



vrije Universiteit *amsterdam*

juli 2014

VOORWOORD

Voor u ligt mijn afstudeerscriptie voor de Master Business Analytics aan de Vrije Universiteit te Amsterdam. Het afstudeeronderzoek heeft plaatsgevonden bij het bedrijf MeasureWorks in Almere in de periode november 2013 tot en met juli 2014. Dit lijkt langer dan er voor gepland staat, maar dat is niet het geval: ik ben vier dagen per week bezig geweest met afstuderen. De vijfde dag ben ik als werknemer in dienst getreden van MeasureWorks.

Tijdens mijn studie ben ik geïnteresseerd geweest in de psychologie. Met deze afstudeeropdracht heb ik de mogelijkheid gehad mij meer in dit onderwerp te verdiepen. Hoewel dit component beperkt moest blijven, ik studeer tenslotte hopelijk toch als MSc. of Business Analytics af, heb ik er veel over gelezen en veel van geleerd.

Mijn dank gaat in de eerste plaats uit naar Raoul Wilmans, mijn begeleider vanuit MeasureWorks. Hij heeft me zeer goed kunnen helpen op momenten dat ik er niet meer uit kwam en heeft met zijn kennis en inzicht een belangrijke rol gespeeld tijdens mijn afstuderen. Daarnaast hebben alle andere collega's binnen MeasureWorks het mede mogelijk gemaakt dat dit onderzoek is voltooid; zij hebben mij altijd geholpen en hebben me veel bijgebracht. Ook hebben ze met mij vele uren aan de tafelvoetbaltafel doorgebracht die wel eens als welkome afleiding heeft dienst heeft gedaan.

Ook Sandjai Bhulai wil ik bedanken voor zijn begeleiding vanuit de Vrije Universiteit en een opoffering van vakantie-uren om mij te kunnen laten afstuderen. Onze maandelijkse gesprekken hebben mij vaak op de goede weg geholpen. Tot slot wil ik de tweede lezer Eduard Belitser bedanken voor het kunnen lezen en beoordelen van mijn scriptie in een zeer kort tijdsbestek.

Lidewij Kooiman

Juli 2014

SAMENVATTING

Het onderzoek bij MeasureWorks heeft zich gericht op het verhogen van de tolerantie van bezoekers ten aanzien van de performance van e-commerce-websites. Performance is hierbij een combinatie van laadtijd en beschikbaarheid van een webpagina en zijn e-commerce-websites waarop klanten producten kunnen aanschaffen. Denk bijvoorbeeld aan websites als Bol.com en Wehkamp.nl. Het doel van deze websites is het optimaliseren van de conversie; de klant op de site laten doen waarvoor het beoogd is, namelijk producten aanschaffen.

Literatuur toont aan dat website-performance invloed heeft op de conversie: een slechte performance leidt tot minder conversie. Hoe is het mogelijk de gevolgen van een zogenoemde performance-degradatie voor de conversie zoveel mogelijk te beperken? Enerzijds bestaat de mogelijkheid de website te optimaliseren zodat de kans op een performance-degradatie wordt geminimaliseerd. Anderzijds is het ook mogelijk de klant te beïnvloeden: wat als de klant toleranter wordt voor de performance en daardoor een (lichte) performance-degradatie accepteert?

De scope van dit onderzoek richt zich op het tweede aspect: het verhogen van de performance-tolerantie. Hiervoor is gekozen omdat MeasureWorks geen directe invloed kan uitoefenen op de structuur en opbouw van een website maar wel over de middelen kan beschikken om de laadtijd-tolerantie mogelijk te kunnen verhogen. Dit laatste zal gebeuren aan de hand van informatievoorziening: wanneer iemand op de hoogte wordt gesteld van een zekere vertraging, neemt de kans dat diegene blijft wachten toe.

In de eerste instantie is er een methodiek ontwikkeld om de performance-degradatie van een website te kunnen signaleren en te bepalen wanneer het noodzakelijk is in te grijpen in de vorm van berichtgeving aan website-bezoekers. Deze methodiek verzamelt haar data via de MeasureWorks-tag die in de te monitoren website moet worden geplaatst. Deze tag is in staat real-time de performance van websites te monitoren. Tevens is de tag in staat een respons van de methodiek te ontvangen om zo output te kunnen genereren op een webpagina in de vorm van een bericht.

De ontwikkeling van de tag, de literatuur en de data-analyse van performance-data van Bol.com tezamen hebben er toe geleid dat de methodiek bestaat uit twee componenten: de ORC_1 - en de ORC_2 -waarde. Hierbij staat ORC voor Online Representative Customer. De eerste waarde is gebaseerd op de navigatiewijze van converterende bezoekers binnen een website. Op basis van deze navigatiewijze is er een zeker onderscheid te maken tussen converterende en niet-converterende bezoekers. De tweede waarde is gebaseerd op de gemiddelde laadtijd van een pagina in combinatie met de lengte van een sessie. Dit laatste is het aantal pageviews van een bezoeker gedurende één websitebezoek. Er is een verband gemeten tussen de beide variabelen. Dit verband wordt in de ORC_2 -waarde samengevat en wanneer deze waarde in een zekere mate wijzigt lijkt er een verandering van websiteperformance te kunnen worden waargenomen.

De combinatie van beide ORC-waarden wordt ingezet om de klant performance-toleranter te maken. Wanneer de kans groot is dat een bezoeker tot conversie over gaat maar hij dreigt af te haken wegens performance-degradatie, wordt er een melding in de webpagina gegenereerd. Deze melding zou er toe moeten leiden dat, ten tijde van performance-degradatie, minder bezoekers de website conversie-loos verlaten.

Ter verificatie en validatie is de MeasureWorks-tag geïmplementeerd in verschillende websites. Het betreft hier echter geen type websites waarvoor het onderzoek is opgesteld: e-commerce websites. Ook gaat het hier om relatief kleine websites, waardoor de ORC-waarden minder betrouwbaar worden.

Het validatieproces van de methodiek van de tag is door het ontbreken van de juiste test-websites niet toereikend. Dit heeft er toe geleid dat er een test ontwikkeld diende te worden om de gevolgen van berichtgeving tijdens performance-degradatie in kaart te brengen. De resultaten laten zien dat er te weinig bewijs is om de verschillende nulhypothese, die stellen dat informatievoorziening tijdens performance-

degradatie geen invloed op de bezoeker heeft, te verwerpen. Een kritische noot hierbij is dat een groot deel van de respondenten achteraf heeft aangegeven de opdracht niet duidelijk te vinden. Dit zorgt ervoor dat er vraagtekens bij de betrouwbaarheid van de resultaten van de test kunnen worden gezet.

De enquête die na de test is ingevuld laat wél zien dat het merendeel van de respondenten informatie wil ontvangen wanneer er sprake is van performance-degradatie op een website. Dit ondersteunt de wetenschappelijke literatuur in Hoofdstuk 2. Hierdoor lijken met name de enquête na de test van essentieel belang te zijn.

Aanbevelingen voor verder onderzoek zijn dan ook het daadwerkelijk implementeren en optimaliseren van de methodiek bij een e-commerce-website. Daarnaast kan de methodiek worden uitgebreid met meerdere ORC-waarden, waarbij valt te denken aan factoren als pagina-object-gegevens, volgorde van het paginabezoek en het door de bezoeker gebruikte operating system.

Op basis van dit onderzoek lijkt het dan ook mogelijk om de performance-tolerantie van bezoekers van e-commerce-websites te verhogen ten tijde van performance-degradatie om zo de conversie te optimaliseren. Het inzetten van de MeasureWorks-tag maakt het mogelijk de website-performance te monitoren. De methodiek die aan de tag gekoppeld bepaald vervolgens op basis van de ORC₁ en ORC₂ wanneer het noodzakelijk is een bericht te sturen. Dit bericht zal dan door de MeasureWorks-tag naar de browser van de desbetreffende bezoeker worden gestuurd.

INHOUD

Voorwoord	i
Samenvatting	ii
Inhoud	iv
1 Inleiding.....	1
1.1 Aanleiding.....	1
1.2 Probleemstelling.....	1
1.3 Onderzoeksvraag	1
1.4 Scope	1
1.5 Onderzoeksmethode	2
1.6 Opbouw rapport	2
2 Literatuur	3
2.1 Online conversie	3
2.2 Website-Performance	5
2.3 Invloed van website-performance op de conversie van e-commerce websites	8
2.4 Verhogen van de conversie door de performance-tolerantie te beïnvloeden.....	11
2.5 Invloed van MeasureWorks op de performance-tolerantie	13
3 De MeasureWorks RUM-tag	14
3.1 Overzicht.....	14
3.2 Implementatie van de tag	15
3.3 Dataverzameling	15
3.4 Het bijhouden van sessies en de cookie-wetgeving	16
4 Data-analyse.....	18
4.1 Dataverzameling	18
4.2 Preparatie final dataset	19
4.3 Analyse Bol-data	22
4.4 conclusie data-analyse	29
5 Methodiek.....	30
5.1 ORC ₁ -waarde: Navigatie door de website	30
5.2 ORC ₂ -waarde: de gemiddelde laadtijd versus de sessielengte	35
6 Implementatie methodiek	44
6.1 Opslag methodiek.....	44
6.2 ORC ₁	44
6.3 ORC ₂	44
6.4 Weergave van de resultaten	44
7 Resultaten	45

7.1	Wolfrijbewijsshop.nl.....	45
8	Test: Hoe reageren bezoekers op berichtgeving bij vertragingen in een webshop?.....	49
8.1	Doel van de test.....	49
8.2	Beschrijving van de test.....	49
8.3	Scenario's.....	50
8.4	Hypothesen	51
8.5	Resultaten.....	52
8.6	Conclusie	57
9	Conclusies en aanbevelingen	58
9.1	Conclusie	58
9.2	Aanbevelingen	59
	Bronnenlijst.....	60
	Appendix	61
A.	PHP CODE.....	61
B.	Herdefiniëring Paginagroepen Bol.com	77

1 INLEIDING

1.1 AANLEIDING

MeasureWorks (MW) is gevestigd in Almere; het bedrijf is gespecialiseerd in het optimaliseren van web-performance en rapporteren hun klanten, bedrijven als ING en Bol.com, over de laadtijden en beschikbaarheid van hun websites en geven aan hoe deze te optimaliseren is voor een verbetering van de omzet. Laadtijd en beschikbaarheid vormen samen de belangrijkste aspecten van performance. De performance is goed wanneer de laadtijd laag is en de beschikbaarheid hoog; de performance is slecht wanneer de laadtijd hoog is en de beschikbaarheid laag.

Het onderzoek bij MeasureWorks heeft zich gericht op het verhogen van de tolerantie van online bezoekers ten aanzien van de performance van e-commerce-websites. E-commerce-websites zijn websites waarop klanten producten kunnen aanschaffen. Denk bijvoorbeeld aan websites als Bol.com en Wehkamp.nl. Het doel van deze websites is het optimaliseren van de conversie; de klant op de site laten doen waarvoor het beoogd is, namelijk producten aanschaffen.

Literatuur toont aan dat de performance invloed heeft op de conversie: een slechte performance leidt tot minder conversie (Forrester, 2009; Gomez, 2010; Velocity, 2009; Radware, 2014). Hoe is het mogelijk de gevolgen van een zogenoemde performance-degradatie voor de conversie zoveel mogelijk te beperken? Enerzijds bestaat de mogelijkheid de website te optimaliseren zodat de kans op een performance-degradatie wordt geminimaliseerd. Anderzijds is het ook mogelijk de klant te beïnvloeden: wat als de klant toleranter wordt voor de performance en daardoor een (lichte) performance-degradatie accepteert?

Voor MeasureWorks is het interessant de laatste optie verder te onderzoeken. Het bedrijf heeft immers geen directe invloed op de achterliggende opbouw en structuur van een website. Wel kan MeasureWorks invloed uitoefenen op de bezoekers van de website middels een zogenaamd 3-th-party element. In dit onderzoek wordt gekeken naar de mogelijkheden die er binnen MeasureWorks zijn om de bezoekers performance-toleranter te maken. Dit leidt tot de probleemstelling in Paragraaf 1.2.

1.2 PROBLEEMSTELLING

MeasureWorks ziet mogelijkheden invloed uit te oefenen op de performance-tolerantie van bezoekers. Hierbij is het voornamelijk interessant om de pijlen te richten op websites waarbij klanten alternatieve websites kunnen bezoeken als ze niet tevreden zijn over de huidige website. Hier komen e-commerce-websites om de hoek kijken. De vraag rijst dan wanneer bezoekers de website verlaten vanwege performance-degradatie en welke middelen er dat moment ingezet moeten worden zodat bezoekers alsnog hun aankopen op de website te blijven doen?

1.3 ONDERZOEKSVRAAG

Hoe is het mogelijk de performance-tolerantie van bezoekers van e-commerce-websites te verhogen ten tijde van performance-degradatie om de conversie te optimaliseren?

1.4 SCOPE

De onderzoeksvraag richt zich op het verhogen van de performance-tolerantie en hoe dit praktisch in is te zetten zodat MeasureWorks een meerwaarde voor haar klanten kan creëren.

In een vroeg stadium van het onderzoek is vastgesteld dat het van groot belang is klanten real-time te kunnen beïnvloeden om de performance-tolerantie te kunnen verhogen. Dit impliceert dat er een manier moet worden gevonden waardoor websitebezoekers real-time beïnvloed kunnen worden. Het is van belang direct te kunnen

ingrijpen ten tijde van performance-degradatie. Dit tweede aspect impliceert dat het te ontwikkelen product ook real-time performance-data moet kunnen vergaren.

1.5 ONDERZOEKSMETHODE

Voor het kunnen beantwoorden van de onderzoeksvraag is literatuuronderzoek noodzakelijk. Daarna dient er kennis over impact van performance op de conversie te worden verworven. Dit heeft plaatsgevonden middels een analyse van de dataset van Bol.com, een klant van MeasureWorks. De derde stap in het onderzoeksproces is het creëren van een methodiek die performance-degradatie tijdig signaleert en de impact van deze degradatie op de bezoeker minimaliseert. Om de methode te valideren en verifiëren zal zij worden getest aan de hand van performance- en conversie data van andere klanten van MeasureWorks.

Dit betekent dat het onderzoek enerzijds een theoretische component bevat (het literatuuronderzoek) en anderzijds, het grotere deel, een praktische component in de vorm van het ontwikkelen en implementeren van de ontwikkelde methodiek.

1.6 OPBOUW RAPPORT

Het rapport is als volgt ingedeeld. Er wordt gestart met een kort literatuuronderzoek in Hoofdstuk 2. Dan volgt er een korte beschrijving van de tijdens het onderzoek ontwikkelde MeasureWorks-tag in Hoofdstuk 3. Vervolgens komt de data analyse uitgebreid aan bod in Hoofdstuk 4 en zal de hierbij vergaarde informatie leiden tot de ontwikkeling van een methodiek in Hoofdstuk 5. Hoofdstuk 6 zal in het teken staan van het toetsen van de wijze waarop de performance-tolerantie mogelijk zal worden verbeterd, waarop hoofdstuk 7 zal volgen met de resultaten van het gehele onderzoek. Vervolgens volgen in Hoofdstuk 8 de conclusie en aanbevelingen. De scriptie zal eindigen met een overzicht van de bronnen in de Bronnenlijst en de Appendix waarin de bijlagen zijn opgenomen.

2 LITERATUUR

Dit hoofdstuk geeft een overzicht van de gebruikte theorie tijdens het onderzoek. Allereerst wordt besproken wat online conversie inhoudt, vervolgens wordt er in gegaan op website-performance en wordt het verband tussen online conversie en website-performance gelegd. Tot slot wordt er in dit hoofdstuk gekeken naar de middelen die kunnen worden ingezet om de negatieve impact van website-performance op conversie te minimaliseren.

2.1 ONLINE CONVERSIE

Een website heeft het doel de bezoeker tot conversie aan te zetten. Conversie, in de breedste zin van het woord, betekent het door iemand laten bereiken van een vooraf vastgesteld doel. Bij e-commerce websites wordt conversie vaak gedefinieerd als 'het omzetten van bezoekers naar betalende klanten' (360°ecommerce, 2014). Echter kent een e-commerce website ook andere conversie-doeleinden. Een voorbeeld hiervan is het verkrijgen van zoveel mogelijk aanmeldingen voor een nieuw opgezette e-mail-campagne. Op andere websites, als bijvoorbeeld blogsites, gelden andere conversiedoeleinden als bijvoorbeeld het lezen van een blog.

2.1.1 CONVERSIE FUNNEL

De conversie-funnel beschrijft het proces dat een website-bezoeker dient te doorlopen voordat hij tot de daadwerkelijke conversie kan overgaan (Webics, 2012). Een voorbeeld van een funnel voor e-commerce websites wordt in het volgende figuur afgebeeld (Figuur 2.1). De getallen in de funnel geven aan welk percentage van de bezoekers doorstroomt naar het volgende niveau van conversie.



Figuur 2.1 – Voorbeeld van een conversie-funnel e-commerce website (Orion Web, 2014).

2.1.2 CONVERSIE-RATIO

De 3% in Figuur 2.1 wordt ook wel de conversie-ratio genoemd. Deze ratio geeft aan welk deel van de bezoekers van een website uiteindelijk tot conversie over gaat. In dit geval is er sprake van conversie wanneer een dienst of product wordt aangeschaft. De conversie-ratio verschilt per website; het gemiddelde van een e-commerce website ligt rond de 2% (MarketingSherpa, 2014).

2.1.3 FACTOREN DIE LEIDEN TOT CONVERSIE

Voor een exacte bepaling van de factoren die leiden tot conversie bestaat er geen eenduidige formule. Wel is het mogelijk de meest belangrijke factoren in kaart te brengen waardoor bezoekers van e-commerce sites al dan niet overgaan tot conversie.

De meest bekende methode om de kans op conversie te bepalen is een heuristiek die is ontwikkeld door het wetenschappelijke onderzoeksbureau 'MarketingExperiments'. Het gaat hier om de op empirisch onderzoek gebaseerde 'Conversion Sequence Heuristic' (MarketingExperiments, 2009). Deze heuristiek is een psychologische benadering voor het bepalen van de kans op conversie en wordt weergegeven in Vergelijking 2.1.

$$C = 4M + 3V + 2(I - F) - 2A$$

met

- C - kans op conversie
- M - de mate van overeenkomst van de verwachting van de klant en het aanbod van de website
- V - helderheid en kracht van aanbod
- I - aansporing om actie te ondernemen
- F - wrijvingsselementen gedurende het proces
- A - vrees tot het invullen van gegevens

Vergelijking 2.1 – Conversion Sequence Heuristic (MarketingExperiments, 2009).

Het is lastig om aan de variabelen in de heuristiek een concrete waarde toe te kennen omdat zij psychologisch van aard zijn. Deze heuristiek geeft aan welke factoren voor de bezoeker een rol spelen bij de mate van conversie en welke verbanden er tussen de factoren bestaan.

Volgens de heuristiek is de grootste factor die van invloed is op de kans op conversie de mate van overeenkomst van de verwachting van de bezoeker en het aanbod van de website. Dit lijkt logisch; wanneer een bezoeker verwacht een specifiek boek op een website te kunnen kopen en dit blijkt niet het geval te zijn, zal de kans op conversie drastisch verminderen. De bezoeker zal waarschijnlijk op zoek gaan naar een andere website die het boek wel aanbiedt. Een bezoeker heeft niet alleen verwachtingen over de producten die worden aangeboden, ook zaken als verzendkosten, afhandelsnelheid, verzendtijd, website-performance spelen mee bij het creëren van de verwachting.

De helderheid en kracht van het aanbod spelen ook een grote rol bij het bepalen van de kans op conversie. Een voorbeeld van een krachtig aanbod is het gratis verzenden van de aangeschafte goederen. Dit zou een meerwaarde kunnen betekenen om het product bij deze aanbieder te kopen en niet bij een vergelijkbare aanbieder die wel verzendkosten rekent.

De aansporing om tot actie over te gaan minus de wrijvingsselementen die tijdens het bestelproces optreden hebben ook effect op de conversie. In welke mate een consument wordt aangespoord om een product te kopen is afhankelijk van bijvoorbeeld e-mailcampagnes. Hoe makkelijker het wordt een product waarvoor (actief) interesse wordt gewekt kan worden aangeschaft, hoe groter de kans dat het product daadwerkelijk wordt aangeschaft. Is het daarentegen ingewikkeld om het product te bestellen, bijvoorbeeld omdat er veel stappen moeten worden genomen alvorens de bestelling plaats kan vinden, dan neemt de kans af dat het bestelproces in zijn geheel wordt voltooid.

De laatste factor die een rol speelt bij de conversie is de vrees tot het invullen van gegevens. Wanneer een website niet betrouwbaar overkomt, zal een bezoeker minder snel geneigd zijn persoonlijke gegevens achter te laten. De vraag rijst dan namelijk wat er met de gegevens wordt gedaan. Men zou kunnen denken dat de gegevens worden gebruikt om bijvoorbeeld periodieke mails te sturen waar de bezoeker niet op zit te wachten.

2.2 WEBSITE-PERFORMANCE

Onder website-performance wordt de combinatie van de beschikbaarheid en laadtijd van de pagina's van een website verstaan. Hierbij geeft de beschikbaarheid aan of een webpagina al dan niet bereikbaar is en geeft de laadtijd van een webpagina aan hoe lang het duurt om deze pagina te laden. Hieronder worden beide onderdelen nader toegelicht.

2.2.1 BESCHIKBAARHEID

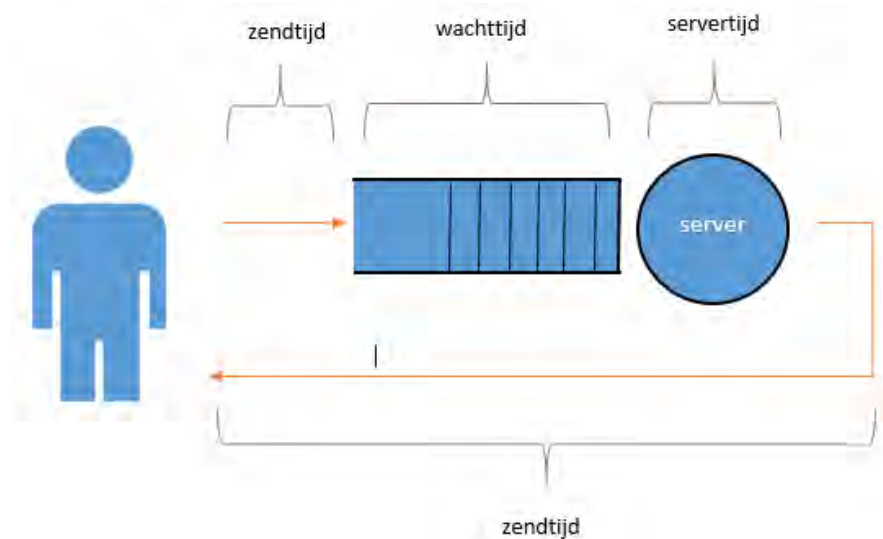
In algemene zin is de definitie van de beschikbaarheid van een webpagina de mate waarop een webpagina bereikbaar en toegankelijk is voor publiek (Seow, 2009). Deze waarde wordt vaak uitgedrukt in een percentage. De beschikbaarheid van een webpagina van afgelopen week kan bijvoorbeeld 99% zijn. In dit geval was de pagina gedurende 99% van deze periode bereikbaar en toegankelijk.

Binnen de beschikbaarheid is er onderscheid te maken tussen de functionele en niet-functionele beschikbaarheid. De functionele beschikbaarheid is het deel van de tijd dat een webpagina in staat is de bezoeker te laten doen waarvoor de website is ontwikkeld. Een website kan in algemene zin beschikbaar zijn en toch niet functioneel beschikbaar zijn. Een voorbeeld is wanneer men producten in een webshop kan bekijken maar het vervolgens niet mogelijk is om het product in het winkelmandje te plaatsen. In dit onderzoek wordt met beschikbaarheid de functionele beschikbaarheid van de webpagina bedoeld.

De niet-functionele beschikbaarheid geeft de mate aan waarin alle objecten in een webpagina worden geladen. Een webpagina kan tegelijkertijd functioneel beschikbaar zijn en niet object-beschikbaar. Dit betekent dat objecten als een plaatje van een banner of een add-in van Facebook niet worden geladen, maar het eigenlijke doel van de website, het aanschaffen van bijvoorbeeld een boek, wel wordt behaald.

2.2.2 LAADTIJD

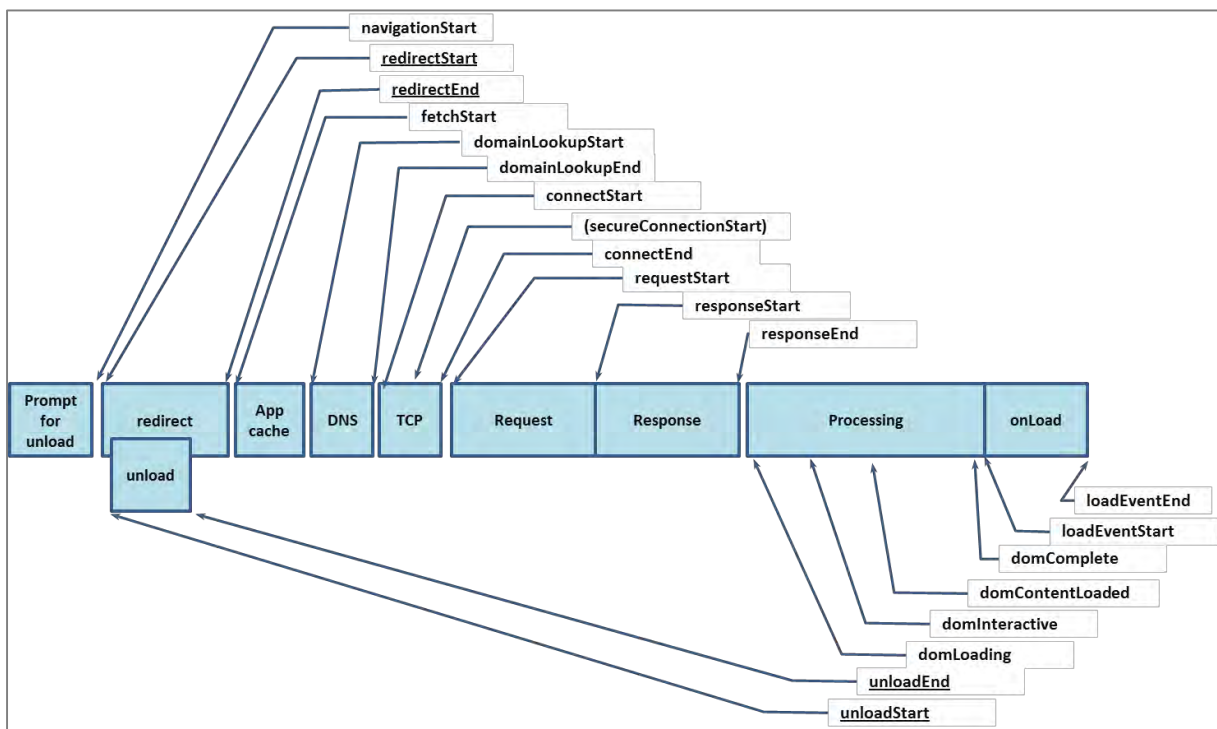
De totale laadtijd van een webpagina is de tijd vanaf het moment dat een pagina wordt opgevraagd, bijvoorbeeld door het klikken op een link, tot aan het moment dat de code achter de pagina volledig is geladen. In andere woorden is de laadtijd de som van de servicetijd, de wachttijd en de zendtijd (Figuur 2.2). De servicetijd is de tijd die het kost om de aanvraag van een pagina te uit te voeren. De wachttijd is de tijd die wordt besteed aan het wachten totdat de aanvraag kan worden behandeld. Tot slot is de zendtijd de tijd die nodig is om de aanvraag via het netwerk te laten reizen.



Figuur 2.2 - Laadtijd opgesplitst in zend-, wacht- en servertijden. De verhoudingen van de procesonderdelen geven niet noodzakelijk een reëel beeld van de werkelijkheid.

Wanneer de laadtijd toeneemt is dit te wijten aan tenminste één van de drie factoren. Niet elke factor heeft evenveel invloed op de variatie van de laadtijd. De servicetijdvariatie is minimaal omdat bij gegeven input een zekere servicetijd geldt. De wachttijd kan wel ongelimiteerd toenemen waardoor de variatie hoog is. Hoe meer aanvragen er bij de webserver worden gedaan, des te sterker de wachttijd toeneemt. Tot slot kent de zendtijd ook een grote variatie. De bandbreedte is hierbij de meeste beperkende factor. Ook het aantal stappen dat moet worden genomen om pakketjes van de server naar de eindgebruiker te sturen (hubs) is bepalend voor de zendtijd.

Een tweede onderscheid binnen de laadtijd kan worden gemaakt door de verschillende technische acties van het laadproces te bekijken. Het is mogelijk om te meten op welke tijdstippen welke acties worden uitgevoerd. Dit zorgt ervoor dat het mogelijk wordt laadtijd aan onderdelen van het laadproces toe te kennen. De tijdstippen waarop de acties plaatsvinden worden in de browser middels de Navigation Timing API vastgesteld. Via deze interface wordt een lijst met verschillende attributen met bijbehorende data opgehaald. Figuur 2.3 hieronder laat zien welke componenten van het laadproces de interface beschikbaar stelt (W3C, 2012).



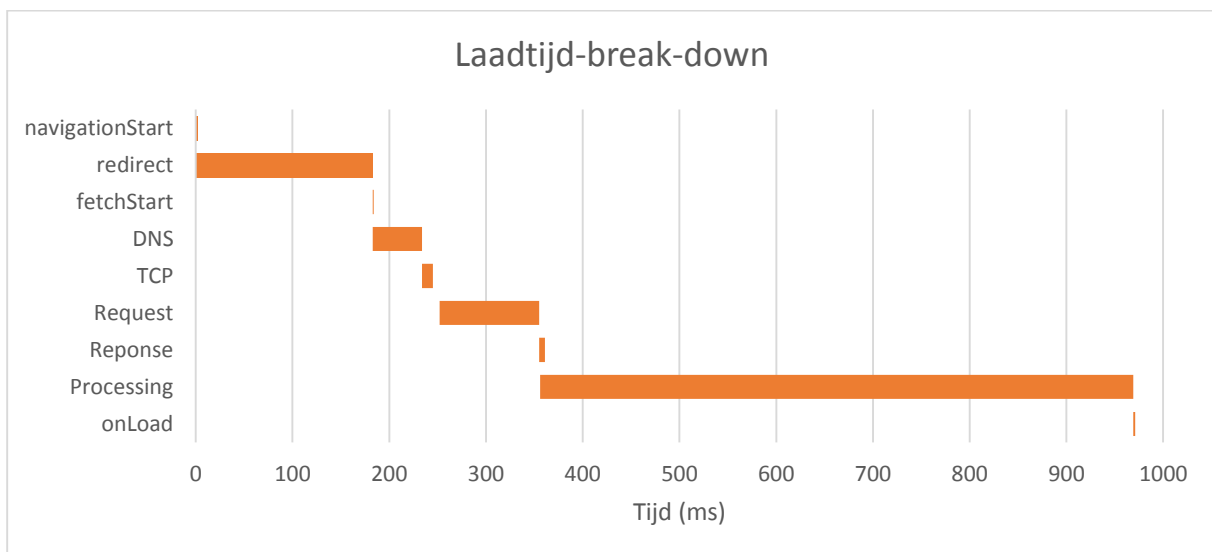
Figuur 2.3 - Navigation Timing API (W3C, 2012).

Vanaf het moment dat een bezoeker navigeert naar een andere pagina treedt het bovenstaande proces in werking. In Tabel 2.1 worden de stappen van dit proces kort toegelicht.

Tabel 2.1 – Laadproces aan de hand van de Navigation Timing API.

Onderdeel laadproces	Toelichting	Meetmoment start	Meetmoment eind
Prompt for unload	START opvragen pagina		navigationStart
Redirect	Tijd die benodigd is om een pagina, indien er een redirect is, door te sturen.	redirectStart	redirectEnd
Unload (tijdens redirect)	Tijd die de browser nodig heeft om de vorige pagina te ontladen. Als de vorige pagina leeg is, is deze tijd 0.	unloadEnd	unloadStart
App cache	De tijd die benodigd is om eventuele onderdelen te cachen	fetchStart	
DNS	De tijd die benodigd is om het IP-adres bij de URL op te vragen.	domainLookupStart	domainLookupEnd
TCP	De tijd die nodig is om een HTTP-connectie op te zetten met de webpagina.	connectStart	connectEnd
Request	Timestamp waarop request aan de webpagina wordt gedaan.	requestStart	
Response	Tijd die benodigd is om een HTTP-response van de webpagina te ontvangen.	responseStart	responseEnd
Processing	Tijd die benodigd is om de HTML-code van de pagina te laden en uit te voeren.	domLoading	domComplete
	*domInteractive en domContentLoaded zijn onderdelen van het onderdeel 'processing'.		
onLoad	Tijd die benodigd is om pagina klaar voor gebruik te maken voor de bezoeker.	loadEventStart	loadEventEnd

Een voorbeeld van Navigation Timing van pageviews van een webpagina is weergegeven in Figuur 2.4.

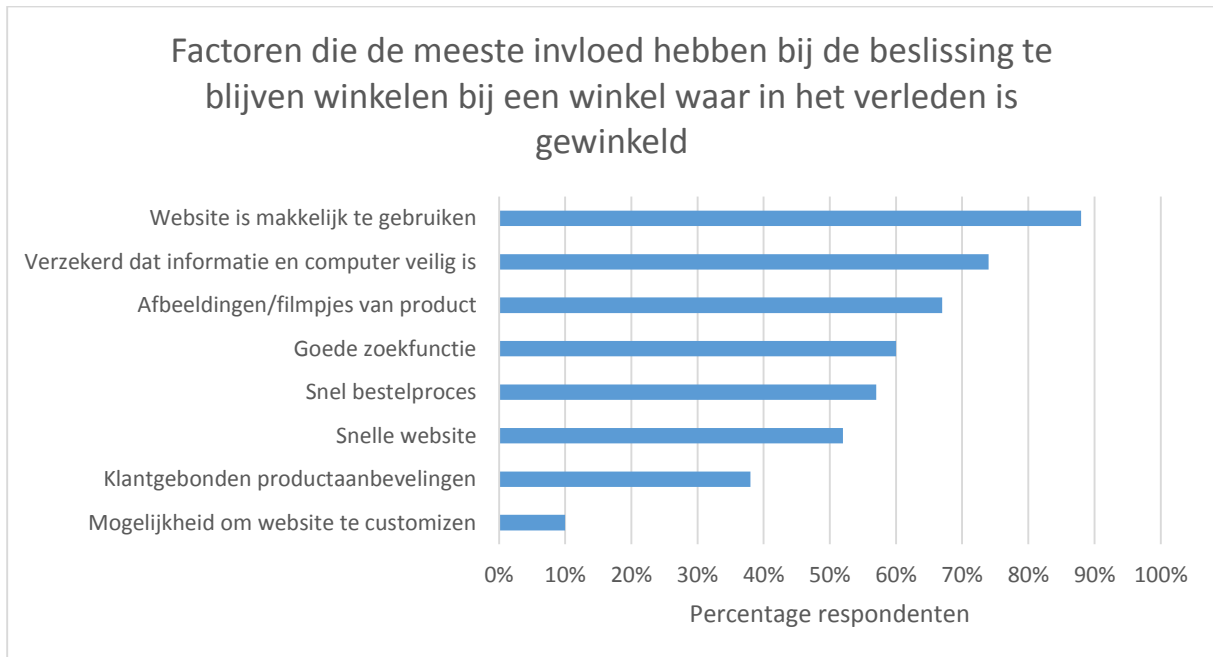


Figuur 2.4 – Voorbeeld laadtijd-break-down aan de hand van de Navigation Timing API.

2.3 INVLOED VAN WEBSITE-PERFORMANCE OP DE CONVERSIE VAN E-COMMERCE WEBSITES

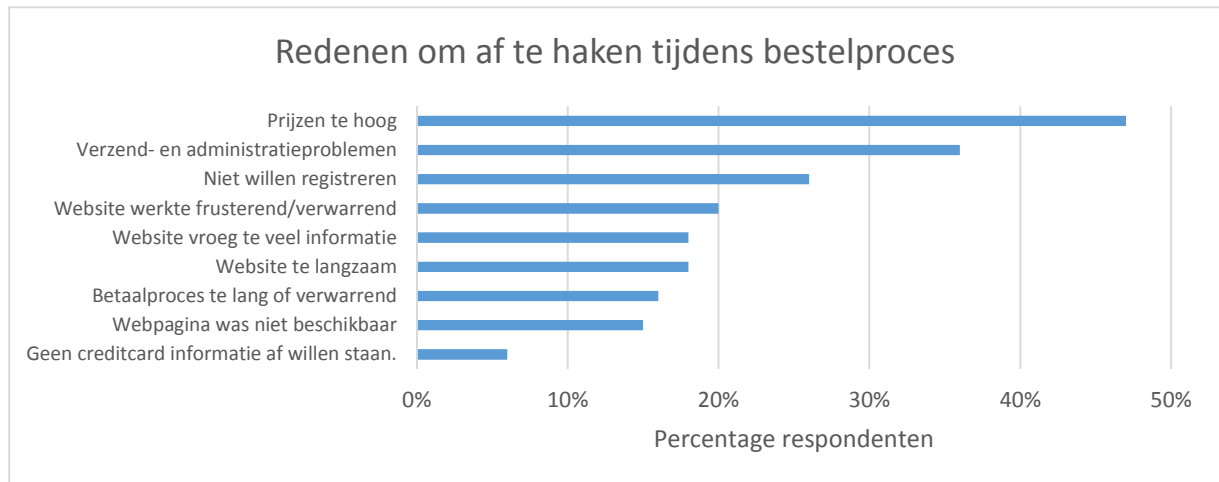
Zoals beschreven in de Paragraaf 2.1, hangt de mate waarin een e-commerce website converteert af van veel factoren. In deze paragraaf gaan we in op de invloed die de performance van een e-commerce website op de conversie heeft.

Uit onderzoek blijkt dat de website-performance van invloed is op de conversie van een e-commerce site. Forrester (2009) heeft onderzoek gedaan naar de redenen dat klanten terugkeren naar een bepaalde webshop. Dit leidde tot de resultaten weergegeven in Figuur 2.5. Meer dan de helft van de respondenten heeft aangegeven dat de snelheid van een website van invloed is bij het maken van de keuze om terug te komen.



Figuur 2.5 - Onderzoek van Forrester (2009) geeft aan wat de redenen zijn die invloed hebben op de keuze van bezoekers om al dan niet terug te keren naar een webshop.

Deze studie geeft ook inzicht in de beweegredenen van bezoekers van online winkels om de website tijdens het daadwerkelijke bestelproces te verlaten. Figuur 2.6 geeft weer waarom klanten die van plan waren een product te kopen, toch zijn afgehaakt. 18% van de ondervraagden geeft aan dat een langzame website een reden is om af te haken tijdens het bestelproces. Ook het niet beschikbaar zijn van een webpagina is voor een bezoeker reden om tijdens het bestelproces af te haken.



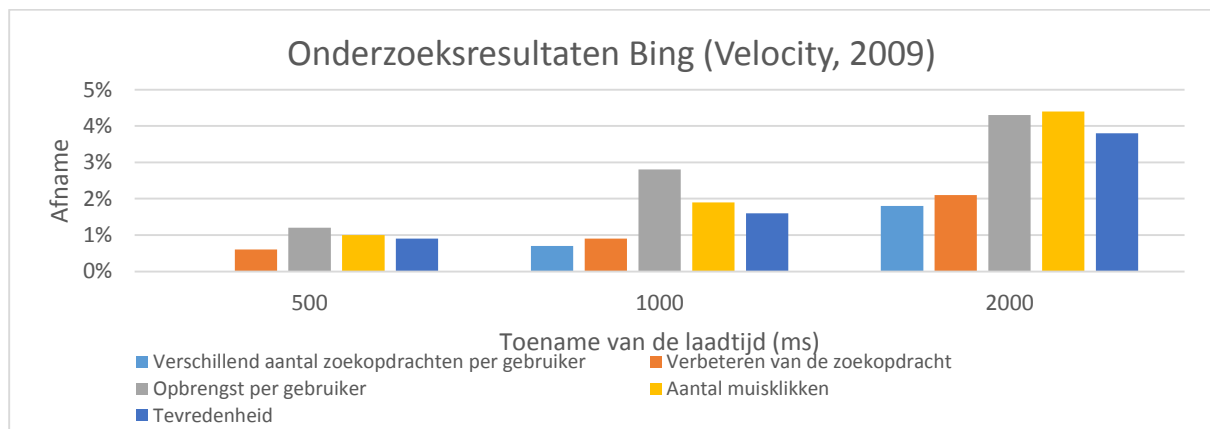
Figuur 2.6 – Onderzoek van Forrester (2009) wijst uit dat een langzame pagina's of pagina's die niet beschikbaar zijn, zorgen dat bezoekers de website verlaten tijdens het bestelproces.

Een studie van Gomez (2010) wijst uit dat 75% van de bezoekers liever de website verlaat voor een concurrent, dan te moeten wachten op vertragingen. Dit is gebaseerd op metingen die hebben plaatsgevonden op een website wanneer het aantal bezoekers zeer hoog was. Ook bleek dat 88% van de bezoekers minder snel terugkomt bij dezelfde e-commerce site na een dergelijke negatieve ervaring bij deze site.

Onderzoek dat werd gepresenteerd op de web-operations en performance conferentie Velocity (2009), liet zien dat een vertraging van twee seconden voor zoekresultaten van Bing leidde tot:

- een afname van 1,8% aan zoektermen
- een afname van 3,75% aan clicks
- meer dan 4% verlies van tevredenheid onder klanten
- 4,3% verlies in opbrengsten per klant

De laadtijd bedroeg in eerste instantie één seconde en door vertragingen te introduceren kon de laadtijd oplopen tot een maximaal drie seconden. Een overzicht van de resultaten van het onderzoek zijn in de onderstaande figuur opgenomen (Figuur 2.7).

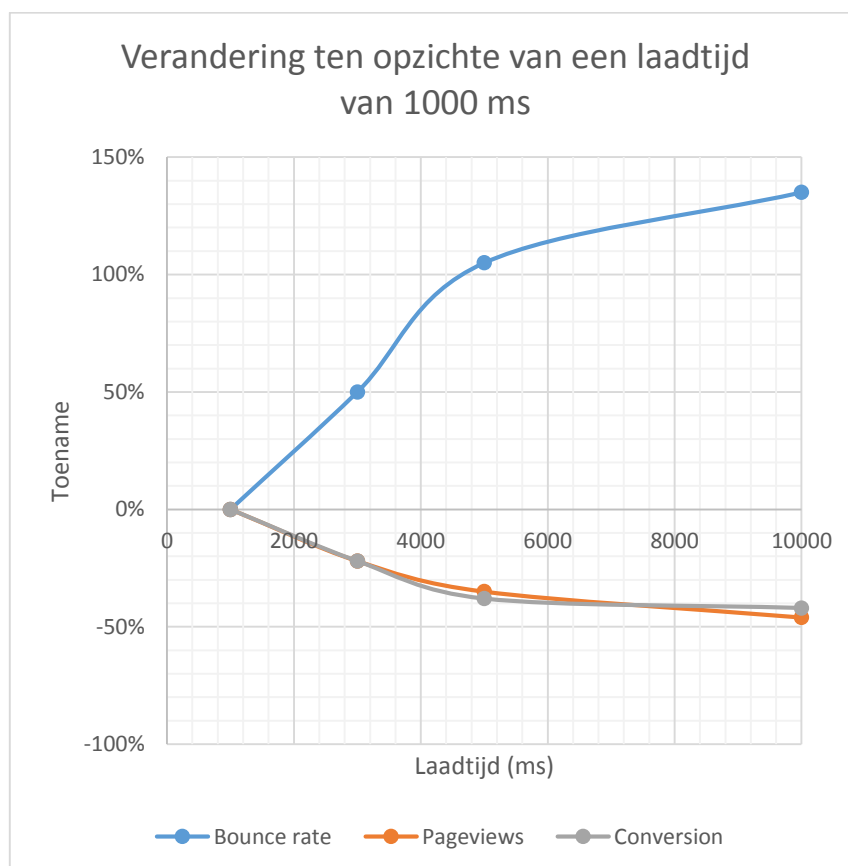


Figuur 2.7 – Onderzoek gepresenteerd op Velocity (2009) wijst uit dat een vertraging op de website leidt tot een lagere conversie.

2.3.1 RADWARE: STATE OF THE UNION

Radware doet elk kwartaal onderzoek naar de performance van de top 500-e-commerce websites in Amerika. De resultaten van dit onderzoek worden kenbaar gemaakt in de 'State of the Union for Ecommerce Web Performance'. Het laatste onderzoek is gepubliceerd in het voorjaar van 2014. Het doel van dit onderzoek is het maken van een momentopname van de mate waarin e-commerce-websites presteren in de echte wereld. Ook geven deze resultaten inzicht in de patronen en trends van website performance.

Het rapport van het voorjaar van 2014 geeft inzicht in de gevolgen van een toename van de laadtijd van e-commerce websites. De resultaten worden in Figuur 2.8 gepresenteerd (Radware, 2014). Hierbij is de bounce-rate het percentage consumenten dat een website bezoekt en deze direct weer verlaat zonder andere pagina's te bezoeken.



Figuur 2.8 – Onderzoek bij zoekmachine Bing wijst uit dat de website-performance invloed heeft op de conversie.

2.4 VERHOGEN VAN DE CONVERSIE DOOR DE PERFORMANCE-TOLERANTIE TE BEÏNVLOEDEN

Uit literatuur blijkt dat website-performance invloed heeft op de conversie; voornamelijk kan worden genoemd dat een hogere laadtijd leidt tot minder conversie. In deze paragraaf wordt uiteengezet hoe de negatieve invloed van de performance op de conversie mogelijk kan worden beperkt. Dit betekent dat er niet wordt getracht de mogelijkheden in kaart te brengen hoe de oorzaak kan worden bestreden (het verbeteren van de performance), maar dat er wordt gekeken naar de mogelijkheden om de gevolgen van een slechte performance in te dammen. Deze aanpak is gekozen omdat MeasureWorks een product wil leveren dat niet afhankelijk is van de opbouw van een website, maar slechts reageert op de gemeten performance.

Het beperken van de negatieve gevolgen van de afname van de performance (hogere laadtijden van pagina's en een afname van de beschikbaarheid) betekent dat bezoekers bij eenzelfde performance anders moeten handelen: klanten dienen een langere wachttijd te tolereren. De vraag is nu hoe deze wachttijd-tolerantie kan worden vergroot. Aan de hand van Maister's First Law wordt getracht aan te geven welke factoren invloed hebben op de wachttijd-tolerantie.

2.4.1 MAISTER'S FIRST LAW

Maister schrijft dat de tevredenheid met betrekking tot een ervaring afhankelijk is van de vooraf gecreëerde verwachting van ervaring en de daadwerkelijke beleving ervan (Maister, 1985). Dit verband staat beter bekend als Maister's First Law en wordt omschreven in Vergelijking 2.2.

$$\text{Tevredenheid} = \text{Perceptie/Ervaring} - \text{Verwachting}$$

Vergelijking 2.2 – Maister's First Law.

Dit betekent dat wanneer je de verwachting van een ervaring vooraf negatief bijstelt, de daadwerkelijke gebeurtenis positiever wordt ervaren. Andersom geldt dit ook: wanneer de verwachting toeneemt, zal men bij een gelijke ervaring minder tevreden zijn over de gebeurtenis.

De regel wordt onderschreven door de factor M in de Conversion Sequence Heuristic zoals beschreven in Paragraaf 2.1. Hierbij is M de mate van overeenkomst van de verwachting van de klant en het aanbod van de website. De factor heeft de meeste impact op de conversie volgens de heuristiek.

De regel van Maister kan ook worden toegepast op de invloed die website-performance op de conversie heeft. Dit betekent dat hoe positiever de laadtijd wordt ervaren en hoe lager de verwachting met betrekking tot de laadtijd, hoe hoger de tevredenheid zal zijn. In termen van laadtijd-tolerantie zou dit betekenen dat hoe beter de ervaring (korte laadtijd) en hoe lager de verwachting (hoge laadtijd), hoe groter de tolerantie is voor de laadtijd. Wanneer de verwachting te laag is zullen klanten in eerste instantie echter minder snel geneigd zijn een webshop te bezoeken. Daarom dient de verwachting hoog te worden gehouden en pas op momenten wanneer de kans vrij groot is er dat er een slechtere-performance-dan-verwacht wordt ervaren, dient er te worden ingegrepen.

Maister beschrijft in zijn onderzoek acht stellingen met betrekking tot de psychologie van wachttijd. Deze stellingen geven een aantal belangrijke inzichten over de perceptie op de wachttijd.

- 1) Bezette wachttijd voelt langer dan onbezette wachttijd. Dit impliceert dat wachten minder erg wordt gevonden als men iets heeft om zijn of haar wachttijd mee te vullen.
- 2) Mensen willen starten met de activiteit. Er heerst meer onzekerheid als de activiteit waarop gewacht wordt nog niet van start is gegaan. Als er eerst gestart wordt en vervolgens nog dient te worden gewacht, ervaart men dit als prettiger dan wanneer de eerste reactie ook op zich laat wachten.

- 3) Angst maakt dat wachten langer lijkt dan daadwerkelijk het geval is. Tijdens het wachten komen er angsten op over de gemaakte keuze.
- 4) Onzekere wachttijd lijkt langer te duren dan bekende eindige wachttijd. Het is minder vervelend om op een afspraak te wachten die over een half uur begint, dan om te wachten op een afspraak die binnen nu en half uur plaatsvindt; de verwachting is dat er een half uur moet worden gewacht. Wanneer de tijd onzeker is, is deze verwachting ook moeilijker in te schatten.
- 5) Niet-onderbouwde wachttijd lijkt langer te duren dan onderbouwde wachttijd. Wanneer de reden van de wachttijd bekend en uitgelegd is, is men bereid langer te wachten.
- 6) Oneerlijke wachttijd wordt als langer ervaren dan eerlijke wachttijd. Als anderen die later aankwamen eerder geholpen worden, wordt de wachttijd als langer ervaren dan wanneer het FIFO-systeem gehanteerd wordt.
- 7) Hoe waardevoller een service, hoe langer iemand bereid is daarop te wachten. Men is langer bereid te wachten bij de dokter dan bij de kassa.
- 8) Alleen wachten wordt als langer ervaren dan het wachten in een groep. Op deze manier wordt je sneller afgeleid waardoor het lijkt alsof de wachttijd korter is. Deze stelling volgt uit de eerste.

Ook Steven Seow (2009) geeft een overzicht van de belangrijkste factoren die tolerantie beïnvloeden. Hij spitst zich hierbij specifiek toe op de website-performance. De belangrijkste factoren zijn:

- *Informatievoorziening*: wanneer een webshop aangeeft een zekere handeling binnen een bepaalde tijd te kunnen uitvoeren en dit niet nakomt, heeft dit een negatief effect op de ervaring van de bezoeker.
- *Betrouwbaarheid en stabiliteit van de website*: hoe vaker het nodig is om langer te wachten of een pagina opnieuw te moeten laden omdat hij het in eerste instantie niet deed, hoe lager de mate van tevredenheid over dit websitebezoek.
- *Emotionele status van de bezoeker*: wanneer iemand een website bezoekt onder negatieve emotionele omstandigheden zal de ervaring negatiever zijn dan bij iemand die positieve emotionele omstandigheden ervaart.
- *Moment in tijd*: op momenten wanneer de urgentie om een website te gebruiken hoog is, leidt een hogere laadtijd vaak tot een slechtere ervaring.
- *Internetgebruik en –ervaring*: een geoefende internetsurfer stelt hogere eisen dan een persoon die weinig van het internet gebruik maakt.
- *Standaarden en benchmarks*: als een gebruiker van een andere webshop een zekere laadtijd gewend is, zal hij sneller verwachten dat een webshop minstens dezelfde performance kent.
- *Eerdere ervaringen op de website*: als de website de vorige keer een zeer goede performance kende, zal de gebruiker verwachten weer een zekere mate van goede performance aan te treffen.
- *Vooringenomenheid*: de mate waarin vooraf (zonder eerdere website-ervaringen) de verwachtingen worden gesteld. Deze verwachtingen worden gebaseerd op bijvoorbeeld de reputatie van een bedrijf.
- *Omgevingsfactoren*: hierbij gaat het om factoren als cultuur, trends en rages.

Tot slot wijst onderzoek (Dellaert, 1999) uit dat consumenten een wachttijd niet noodzakelijk als negatief bestempelen wanneer ze terugdenken aan een websitebezoek waarbij de wachttijd goed gemanaged is. Dit houdt in dat wanneer er juiste en voldoende informatie met betrekking tot de wachttijd wordt verstrekt, de verwachtingen worden bijgesteld. Ook laat dit onderzoek zien dat wanneer dat verwachte laadtijd hoger ligt dan de ervaren laadtijd, de gevolgen van de laadtijd op de ervaring van de websitebezoeker minimaal zijn.

2.5 INVLOED VAN MEASUREWORKS OP DE PERFORMANCE-TOLERANTIE

In de vorige paragraaf zijn de factoren beschreven die invloed uitoefenen op de performance-tolerantie. In deze paragraaf wordt gekeken naar de mogelijkheden die MeasureWorks als derde partij heeft om de performance-tolerantie van bezoekers te verhogen.

2.5.1 KEUZE VOOR FACTOR OM DE PERFORMANCE-TOLERANTIE TE VERHOGEN

MeasureWorks is een derde partij en is daarom beperkt in haar mogelijkheden de eerder beschreven factoren omtrent de tolerantie te beïnvloeden. Bij de keuze van de te beïnvloeden factoren wordt daarom rekening gehouden met de volgende aspecten:

- Mogelijkheid tot beïnvloeden: op welke factoren is door MeasureWorks de meeste invloed uit te oefenen?
- Mogelijkheid om een product te lanceren: met welke factoren is het mogelijk een product te ontwikkelen dat bij meerdere klanten kan worden ingezet?
- Conversie verhogen: welke factoren dragen bij aan het optimaliseren van de conversie van een webshop?

Op basis van deze componenten en de gegeven argumenten in de vorige paragrafen is gekozen de factor informatievoorziening te beïnvloeden. Omdat meerdere auteurs het nut van informatievoorziening tijdens wachttijd onderschrijven, lijkt het informeren van bezoekers ten tijden van performance-degradatie de negatieve effecten van de performance-degradatie te verlagen. Op deze manier wordt er meer zekerheid over het wachtproces gecreëerd waardoor volgens de literatuur de wachttijd-tolerantie worden verhoogd.

2.5.2 WIJZE VAN BERICHTGEVING

Er zijn verschillende manieren om de klant informatie te tonen over het wachtproces. Voorbeelden hiervan zijn het incrementeel laden van een webpagina of concrete berichtgeving over de performance.

Onderzoek (Bhatti, 2000) toont aan dat wanneer een webpagina incrementeel wordt geladen, dit wil zeggen dat er stukje bij beetje meer verschijnt op een webpagina, gebruikers bereid zijn langer te wachten totdat de pagina geladen is. Tot een laadtijd van 39 seconden beoordeelden gebruikers de kwaliteit van de service als hoog. Dit in tegenstelling tot wanneer een pagina in één keer werd geladen: de hoogste laadtijd waarbij de kwaliteit van de service als hoog werd bestempeld bedroeg 5 seconden.

Ook bleek uit dit onderzoek dat wanneer een verzoek niet binnen vijf tot tien seconden werd voltooid, er een indicatie zou moeten worden gestuurd dat de website nog bezig is met laden, en dus niet is vastgelopen. Op deze manier verzeker je de bezoeker ervan dat het proces nog gaande is en er resultaat verwacht kan worden. Wanneer bezoekers niet op de hoogte worden gesteld van de vertraging is de performance-tolerantie aanzienlijk lager en het percentage afhakers groter.

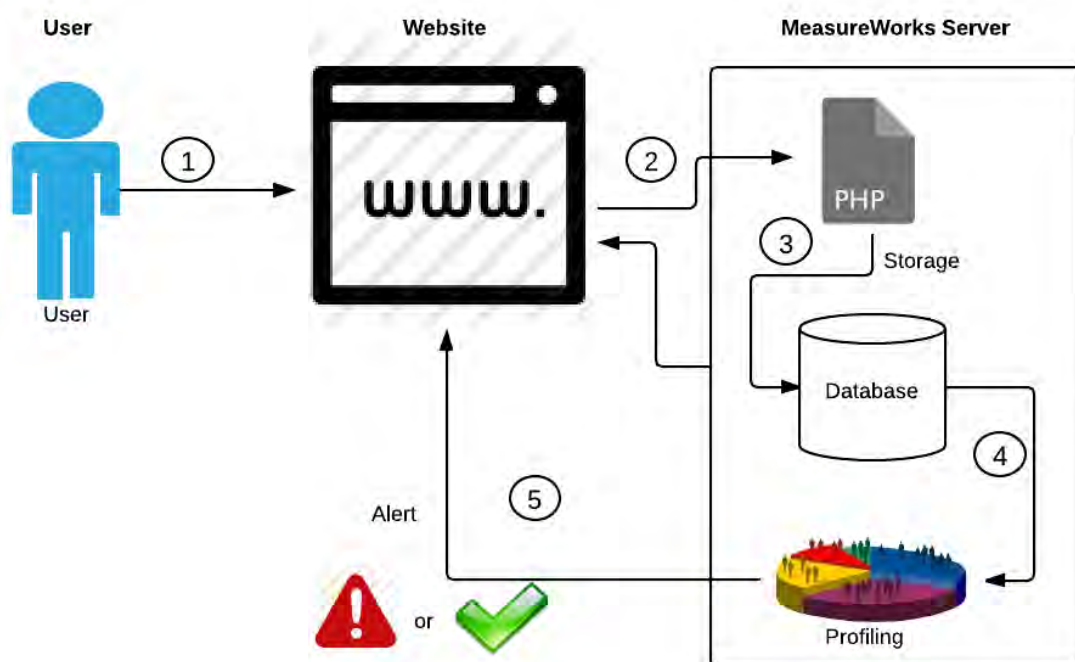
Omdat MeasureWorks geen directe invloed heeft op het incrementeel laden van een webpagina, zal het product zich baseren op het sturen van concrete performance-informatie naar de gebruiker. Dit betekent dat de bezoeker ten tijde van performance-degradatie wordt gemeld dat deze waarschijnlijk een slechtere performance dan gewend, kan verwachten.

3 DE MEASUREWORKS RUM-TAG

Dit hoofdstuk beschrijft de wijze waarop het technisch gezien mogelijk wordt om bezoekers een bericht te sturen ten tijde van performance-degradatie. Dit gebeurt via de MeasureWorks RUM-tag (MW-tag). RUM staat voor Real User Monitoring en houdt in dat elke pageview van een bezoeker kan worden gemonitord. Middels de RUM-tag wordt informatie naar MeasureWorks (MW) gestuurd alwaar deze data gemonitord wordt. Naar aanleiding daarvan kan, indien nodig, informatie naar de gebruiker worden teruggestuurd. In dit hoofdstuk wordt de MeasureWorks RUM-tag beschreven.

3.1 OVERZICHT

De onderstaande figuur geeft inzicht in de werking van de MeasureWorks tag. Het proces bestaat uit een vijftal stappen die onder het figuur worden beschreven (Figuur 3.1).



Figuur 3.1 - Werking MeasureWorks RUM-tag.

- 1) Een bezoeker bezoekt een pagina van een website waarop de tag van MeasureWorks is geplaatst. Tijdens het laden van de pagina wordt de MW-tag in de vorm van een stuk JavaScript uitgevoerd. Hierbij worden gegevens als laadtijd, browser type en sessie-ID uit de browser uitgelezen.
- 2) De door de tag verzamelde informatie wordt naar de server van MW gestuurd en opgevangen door een PHP-script.
- 3) De binnengekomen data wordt vervolgens opgeslagen in een database.
- 4) De methodiek met betrekking tot de performance-tolerantie wordt ook middels het in 2) genoemde PHP-script uitgevoerd.
- 5) Het PHP-script ontvangt van het PHP-script een response. Dit kan ofwel betekenen dat er niets hoeft te worden gedaan, ofwel kan het betekenen dat er middels het JavaScript (uit stap 1) een melding wordt gegenereerd op de websitepagina.

3.2 IMPLEMENTATIE VAN DE TAG

De tag is een stuk JavaScript dat de juiste data naar de servers van MeasureWorks kan sturen én een mogelijke notificatie voor het tonen van een melding kan ontvangen. In het stukje code kan men een label, ook wel paginagroep genoemd, aangeven die een karakterisering van de betreffende pagina geeft op basis van het conversieproces. Dit zijn dus labels als 'home', 'zoeken', 'productcategorieën', 'producten', 'winkelmandje', 'bestellen' en 'besteld'. Een voorbeeld van een webpagina die onder het label 'producten' valt, is de pagina waarop een specifiek item kan worden besteld. Pagina's die in de paginagroep 'zoeken' zouden vallen zijn de pagina's met de zoekresultaten.

Er zijn pagina's die als enige tot een specifiek label behoren. Een voorbeeld hiervan is de pagina van het winkelmandje. De reden hiervoor is dat deze pagina een unieke stap is in het conversie funnel.

3.3 DATAVERZAMELING

3.3.1 NAVIGATION TIMING

De data die de tag verzameld wordt grotendeels opgevraagd via de Navigation Timing API zoals in Paragraaf 2.2 nader is toegelicht. Deze API wordt niet door iedere browser ondersteund. Dit betekent dat de tag niet alle bezoeken kan registreren. De browsers die de API wel ondersteunen worden in de onderstaande tabel (Tabel 3.1) weergegeven en zijn gebaseerd op cijfers uit 2013 (StatsCounter, 2014). Ook is terug te vinden welk percentage van de desktop-gebruikers gebruik maakt van de browser (Can I Use, 2014). In de tabel is te zien dat 80,3% van de op-een-desktop-gebruikte-browsers Navigation Timing ondersteunt in 2013.

Tabel 3.1 – Desktop-browsers die Navigation Timing ondersteunen (StatsCounter, 2014).

Browser	Ondersteund vanaf versie	% pageviews op desktop in 2013
Chrome	6.0	36,1%
Internet Explorer	9.0	27,7%
Firefox	7.0	16,4%
Opera	15.0	0,1%

3.3.2 OVERIGE DATA

Naast het verzamelen van de performance data worden er nog een aantal waarden opgeslagen. Deze worden weergegeven in Tabel 3.2.

Tabel 3.2 – Data die naaste de performance-data beschikbaar wordt gesteld door de Navigation Timing API.

Variabele	Toelichting
Tag-ID	Uniek ID van de pageview.
Invoeg tijdstip	Moment waarop de pageview daadwerkelijk in de database wordt opgeslagen.
URL	URL van de pageview.
Sessie-ID	Uniek sessie-ID per sessie.
Versie MW-tag	De versie van de MW-tag waarmee de pageview wordt geregistreerd.
Pagelabel	Paginalabel die per pagina(groep) door de website-beheerder is toegekend.
User Agent	De user agent geeft aan welke browser en welke browserversie gebruikt zijn tijdens het bezoek aan de webpagina.
Referrer	De URL van de vorige webpagina die door de gebruiker is bezocht.

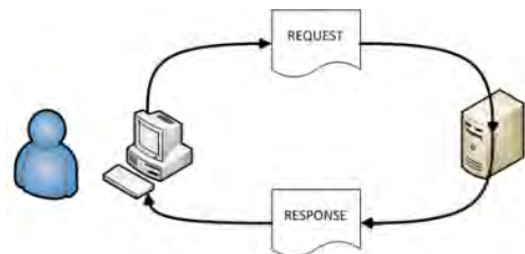
3.4 HET BIJHOUDEN VAN SESSIES EN DE COOKIE-WETGEVING

Om de consument zo goed mogelijk van feedback te kunnen voorzien is het noodzakelijk bij te houden welke pagina's deze consument bezoekt. Dit betekent dat er sessiegegevens bijgehouden moeten worden. Omdat het bijhouden van gegevens veel vraagtekens oproept met betrekking tot de privacy, is het van belang dat de wijze waarop de MeasureWorks-tag de data collecteert en opslaat niet in strijd is met de huidige wet- en regelgeving. In deze paragraaf wordt beschreven op welke manier de data wordt verzameld en opgeslagen.

3.4.1 COOKIES

De meest gebruikte manier om bezoekersgegevens op te slaan is het gebruiken van cookies. Cookies zijn hoeveelheden data die door een website op de computer van de gebruiker worden geplaatst om informatie te bewaren. Websites gebruiken de cookies vervolgens om specifieke gebruikersgegevens op te kunnen slaan. Aan de hand van deze data is het mogelijk dat een bezoeker ingelogd blijft wanneer hij naar een volgende pagina navigeert. Daarnaast kan een website data verzamelen om specifieke producten aan te raden of persoonlijke berichten te sturen.

Er zijn twee verschillende soorten cookies, te weten de response-cookies en de request-cookies. De response-cookies zijn de cookies die door de browser naar de computer van de bezoeker worden gestuurd. De request-cookies zijn de cookies die vanaf de computer van de gebruiker worden meegestuurd wanneer deze een aanvraag doet voor een webpagina. De twee typen cookies worden weergegeven in Figuur 3.2.

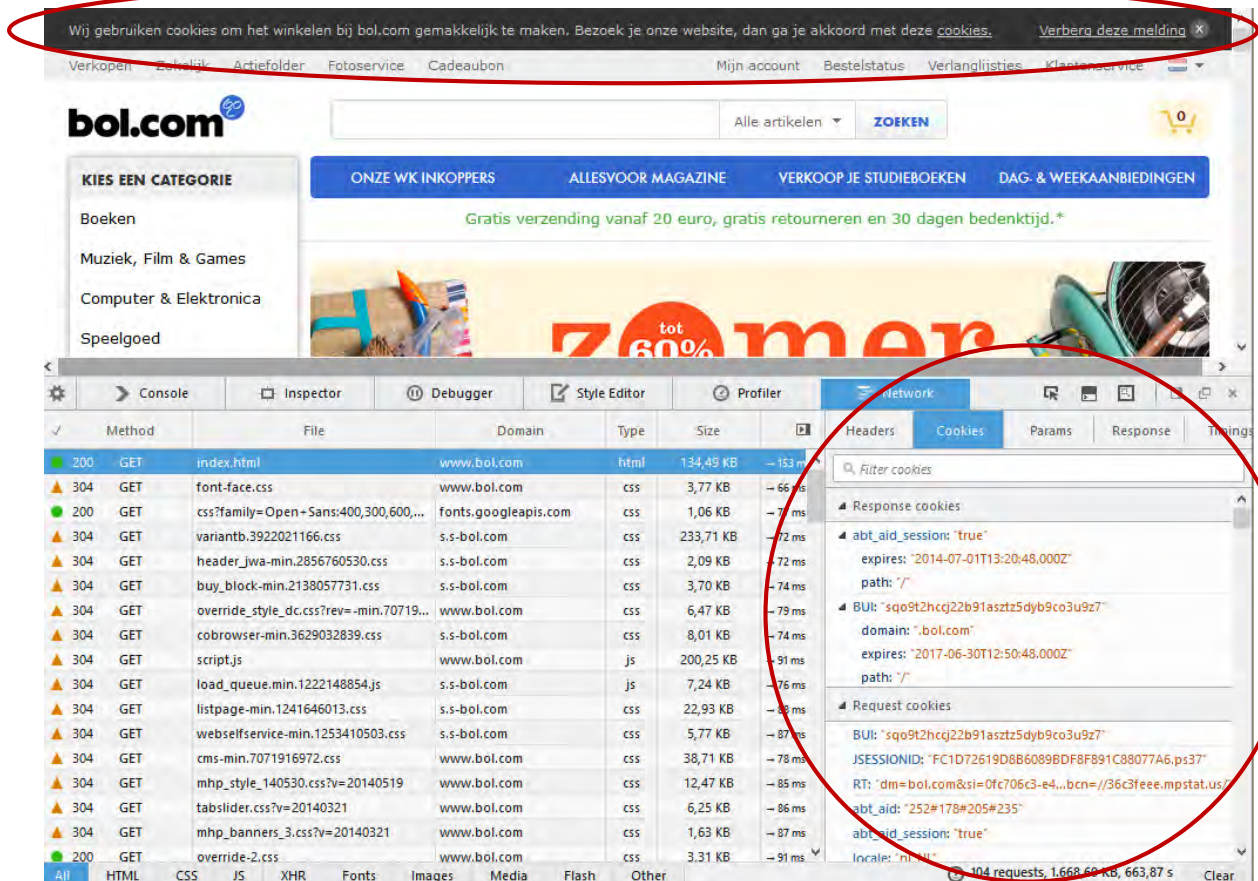


Figuur 3.2 – Onderscheid response- en request-cookies

In de browser zijn de cookie-gegevens door de bezoeker op te vragen. Een voorbeeld van hoe dit in de browser (Firefox versie 30.0) wordt weergegeven is te vinden in Figuur 3.3. Hierbij gaat het om een cookie die door Bol.com is geplaatst. De figuur is gebaseerd op een bezoek aan bol.com dat dateert van 1 juli 2014 14.50u.

De geldigheid van verschillende cookies varieert. Te zien is dat de eerste response-cookies een half uur geldig blijft, de tweede cookie blijft daarentegen 3 jaar geldig. De request-cookies kennen geen houdbaarheid omdat deze zijn gebaseerd op de reeds geplaatste cookies op de computer van de gebruiker. Dit betekent dat wanneer de houdbaarheidsdatum van deze cookies overschreden wordt er automatisch een nieuwe cookie wordt geplaatst middels de request-cookies.

De laatste jaren is er veel ophef ontstaan over het gebruik van cookies door websites. Tot juni 2012 kon iedere website ongevraagd cookies plaatsen. Vanaf 5 juni 2012 trad Artikel 11.7a van de Telecommunicatiewet (in de volksmond "de Cookiewet" genoemd) in werking. Vanaf dat moment moesten websites toestemming aan de bezoeker vragen om cookies te plaatsen. Ook dienen zij de bezoeker informatie te verschaffen over welke informatie er aan de hand van de cookies wordt opgeslagen. Dit is te zien bovenin Figuur 3.3.



Figuur 3.3 - Cookies op een webpagina van Bol.com.

3.4.2 BROWSER WINDOW OBJECT

Wanneer MeasureWorks ook cookies zou plaatsen op een website, zou dit betekenen dat websites die de tag implementeren deze tag ook in de cookie-informatievoorziening zouden moeten opnemen. In dat geval wordt het proces om de tag in gebruik te nemen een stuk complexer. Daarom is er voor gekozen een andere oplossing te gebruiken om de performance-gegevens op te kunnen slaan. Omdat de tag geen persoonsgegevens verzamelt, valt deze oplossing binnen de toegestane wet- en regelgeving.

Het kunnen koppelen van de verschillende pageviews aan één sessie geschiedt daarom middels het overschrijven van een eigenschap van het window-object in een webpagina. Elke tab van de browser kent zijn eigen window-object en elk window-object heeft de eigenschap 'name'. Normaal werd deze eigenschap gebruikt om communicatie tussen verschillende pop-up-windows mogelijk te maken. Tegenwoordig wordt er nauwelijks gebruik gemaakt van pop-up-windows. Dit betekent dat de naam van het window-object in de meeste gevallen niet door een website wordt toegekend. Bij het ontwikkelen van de tag is van dit gegeven gebruik gemaakt: om de verschillende pagina's binnen één sessie aan elkaar te koppelen wordt bij de eerste pagina van het website-bezoek de naam van het window-object aangepast naar sessie-ID. Wanneer deze naam tijdens de sessie niet door de website wordt aangepast, is het voor de tag mogelijk om een gebruiker te blijven volgen. Zodra het tabblad wordt afgesloten, verdwijnt ook het door de tag toegekende sessie-ID.

De gevolgen van het volgen van sessies via het window-object zijn onder andere dat er per tabblad een sessie kan worden bijgehouden. Wanneer men via meerdere tabbladen tegelijk een website bezoekt, kunnen deze twee bezoeken niet aan elkaar worden gekoppeld. Wanneer een bezoeker namelijk een nieuw tabblad opent voor een product, krijgt dit tabblad een nieuw sessie-ID toegekend. Een ander gevolg van de wijze van volgen middels het window-object is dat het sessie-ID wordt overschreven wanneer de naam van het window-object wel is gedefinieerd.

4 DATA-ANALYSE

4.1 DATAVERZAMELING

Om de methodiek voor de tag te kunnen ontwikkelen is er een dataset nodig die vergelijkbaar is met de data die de tag verzamelt. Omdat de tag pas in een later stadium van het project gereed is, is er een andere databron nodig. MeasureWorks beschikt over data van Bol.com waarvan het type grotendeels overeenkomt met het type data dat middels de tag wordt opgeslagen. De data-analyse heeft daarom plaatsgevonden op deze data.

4.1.1 DATASET BOL.COM

De data van het domein van Bol.com is door MeasureWorks verkregen via een tag van een externe partij: SOASTA Real User Monitoring tag (RUM-tag). De SOASTA-tag wordt, net als de MW-tag, in de broncode van een webpagina geplaatst. Vervolgens worden de geregistreerde gegevens naar de databases van SOASTA gestuurd. MeasureWorks haalt de gegevens vervolgens op via een interface en plaatst ze in haar eigen datawarehouse (DWH).

4.1.2 SOASTA

Bol.com heeft de tag van SOASTA in de broncode van haar webpagina's geplaatst. Deze tag stuurt gegevens als laadtijden en URL door naar SOASTA. Ook stuurt de tag de gegevens van unieke sessies door: Bol.com kan door het gebruik van cookies bijhouden hoe een bezoeker door de website heen klikt; elke bezoeker krijgt dan een eigen sessie-ID die een bepaalde periode geldig blijft. Hierbij is het wel van belang dat de gebruiker de cookies accepteert; als hij dit niet doet kunnen zijn gegevens niet worden bijgehouden. Cookies moeten per browser geaccepteerd worden: als een bezoeker de website eerst via Chrome bezoekt en daarna via Firefox krijgt hij twee verschillende sessie-ID's toegewezen. Als de bezoeker zijn of haar cookies tijdens zijn bezoek wist, gaat dit sessie-ID echter verloren. Meer informatie over cookies is te vinden in Hoofdstuk 3.

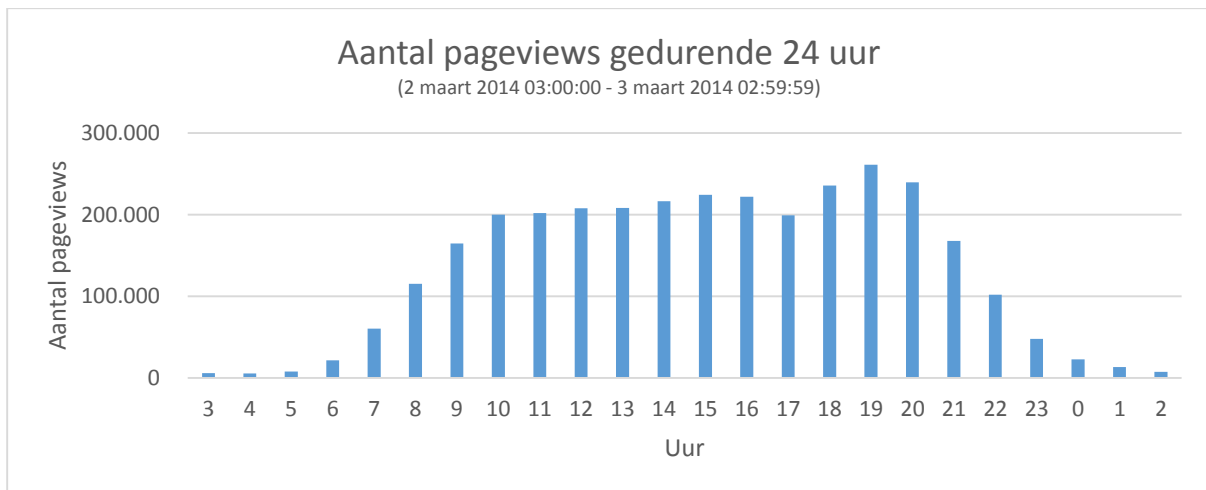
De SOASTA-tag maakt net als de MeasureWorks-tag gebruik van paginalabels. Elke pagina heeft hierbij een label dat niet uniek hoeft te zijn maar het mogelijk maakt pagina's te groeperen.

4.1.3 BEPERKINGEN VAN DE DATA

Na analyse van de eerste versie van de dataset van Bol.com bleek dat de sessie-ID's niet juist werden opgeslagen: op het moment dat een gebruiker van een niet-beveiligde verbinding naar een beveiligde verbinding overging, veranderde het sessie-ID van deze gebruiker. Op deze manier was het niet mogelijk een gehele sessie van een gebruiker te volgen. Dit probleem ontstond bij het binnenhalen van de data door MeasureWorks. Dit is vervolgens opgelost waardoor de tweede versie van de data de juiste sessie-ID's bevatte. Helaas heeft de data-analyse hierdoor later plaatsgevonden dan gepland.

4.1.4 EIGENSCHAPPEN VAN DE DATASET

De laatste versie van de data bevat één etmaal aan pageviews van de webwinkel Bol.com. Dit is het etmaal vanaf 2 maart 2014 03:00:00 tot en met 3 maart 2014 02:59:59. De reden dat voor deze periode is gekozen, is dat uit de eerste versie van de dataset bleek dat er het aantal pageviews rond drie uur 's nachts het laagst is. Hierdoor is de kans dat je 'inbreekt' in sessies het kleinst. De keuze om een dag aan data te gebruiken is gebaseerd op de hoeveelheid data. Het aantal pageviews dat één dag aan data bevat is ongeveer 3 miljoen. De grafiek hieronder laat dit zien (Figuur 4.1).



Figuur 4.1 – Aantal pageviews verdeeld over het uur van de dag.

4.1.5 OPSLAG VAN DE DATA

Omdat er een aantal miljoen pageviews per dag worden gegenereerd in de webwinkel van Bol.com is er voor gekozen de data in een MySQL-database op te slaan. De aangeleverde data omvat de variabelen weergegeven in Tabel 4.1.

Tabel 4.1 - Variabelen in de dataset.

Variabelen	Type	Uitleg	Beschrijving waarden
ID	Integer	Elke pageview kent zijn eigen ID	0 - 3159508
timestamp	Datum + tijd	Dit is van het type jaar-maand-dag uur: minuut: seconde	2014-03-02 03:00 - 2014-03-02 02:59:59
tag_session_ID	Integer	Elke nieuwe gebruiker krijgt een uniek sessie-ID per sessie.	115928019 - 211430714
page_RUM_label	Tekst	Bol.com heeft zijn URL's onderverdeeld in RUM-labels. Het aantal RUM-labels varieert en ligt rond de 200	192 verschillende teksten
page_RUM_label_ID	Integer	Het ID van het RUM-label	89 - 7354
OS_beschrijving	Tekst	Het type van het operating system	26 verschillende teksten
OS_ID	Integer	Het ID van het type operating systeem	7 - 276
URL_host	Tekst	De URL van pageview	846420 verschillende teksten
URL_host_ID	Integer	Het URL-ID van pageview	2146700 - 102271482
domComplete_ms	Integer	Tijd in ms van muisklik totdat DOM compleet geladen is	0 - 119688
loadEventEnd_ms	Integer	Tijd in ms van muisklik tot het Load Event End	0 - 119998
IP_adres	Tekst	Het IP-adres van de computer van de bezoeker	340572
land	Tekst	Het land waar het request wordt gedaan	153 verschillende teksten
geoGroep_ID	Integer	Het ID van de plaats waar het request is gedaan	3 - 925485

4.2 PREPARATIE FINAL DATASET

In deze paragraaf wordt de dataset geprepareerd voor verdere analyse: welke beperkingen kent de data en op welke wijze dient de data gefilterd te worden. Het einde van deze paragraaf levert een data set die als input wordt gebruikt voor het ontwikkelen van de methodiek.

4.2.1 DATA-FILTERING

Niet alle pageviews zijn interessant voor het onderzoek. Zo kan er onderscheid gemaakt worden tussen pageviews die op een desktop worden gemaakt en pageviews die via een mobiel apparaat worden gemaakt. Ook andere factoren als het land waarin de pageview wordt gemaakt en het gebruikte operating systeem worden onder de loep genomen.

4.2.1.1 DESKTOP VERSUS MOBIELE WEBSITE

Onder de aanname dat mobiele website-gebruikers zich op het web anders gedragen dan desktop-website-gebruikers, is besloten het onderzoek in de eerste instantie te richten op de desktop-website. Dit betekent dat alle sessies die een mobiele URL bevatten uit de data zijn gefilterd. Het gaat hierbij om de pageviews met URL's die de string '://m.bol.com' bevatten. Desktop-bezoeken gaan via de URL waarin '://www.bol.com' voorkomt. Tabel 4.2 geeft een overzicht van de data op basis van het onderscheid naar desktop en mobiele gebruiker.

Tabel 4.2 - Onderscheid mobiele en desktop pageviews.

	Desktop website	Mobiele website	Totaal
Aantal pageviews	2.942.961	215.807	3.158.768
Aantal sessies	452.163	56.462	508.625
Gemiddeld aantal pageviews per sessie	6,51	3,82	6,21

4.2.1.2 OPERATING SYSTEM

Omdat de scope van het onderzoek zich richt op desktop-gebruik van de website zijn de overige varianten van gebruik uit de data verwijderd. Het gaat hierbij om de mobiele systemen en de systemen die in zeer kleine mate worden gebruikt. De meeste sessies op mobiele apparaten zijn er al uitgefilterd door de mobiele URL's er uit te filteren. Deze sessies zijn echter nog wel opgenomen in de tabel van de aantallen per operating system (Tabel 4.3).

Tabel 4.3 – Onderscheid per operating system.

Behouden		Verwijderd	
Operating system	Aantal sessies	Operating system	Aantal sessies
Windows 7	200.439	iOS	48.881
Windows	40.111	Android OS	36.948
Mac OS X	34.163	Chrome OS	768
Windows Vista	32.164	Windows Phone	201
Windows 8	27.628	Windows RT	158
Windows XP	20.325	Blackberry OS	123
Ubuntu	1.585	Blackberry Tablet OS	86
Linux	1.501	Windows 2000	14
		Windows NT 4.0	13
		Fedora	5
		SUSE	7
		webOS	4
		Solaris	2
		FreeBSD	2
		Linux Mint	2
		Windows 98	1
		Symbian^3 Belle	1
		Symbian^3 Anna	1
		Undefined	232
Totaal	357.916	Totaal	87.449

4.2.1.3 LANDEN

Bol.com opereert voornamelijk in Nederland en België. Daarom is er voor gekozen de overige landen er uit te filteren. Ook vanwege tijdsinvloeden is het betrouwbaarder om de overige landen niet mee te nemen in de data analyse (Tabel 4.4).

Tabel 4.4 - Onderscheid per land.

Landen	Aantal sessies	Aandeel in totaal
Nederland	373.463	84,5%
België	66.814	15,1%
Overige landen	1.927	0,4%
Totaal	442.204	100%

4.2.1.4 LOADEVENTEND

De performance data die in de dataset staat heeft betrekking op de tijdsmeting op het moment van het LoadEventEnd en domComplete. Omdat er in de eerste instantie gekeken zal worden naar het moment LoadEventEnd worden de sessies met pageviews die daar "NULL"-waardes bevatten verwijderd uit de data. Het gaat hierbij om 9 sessies en 34 pageviews.

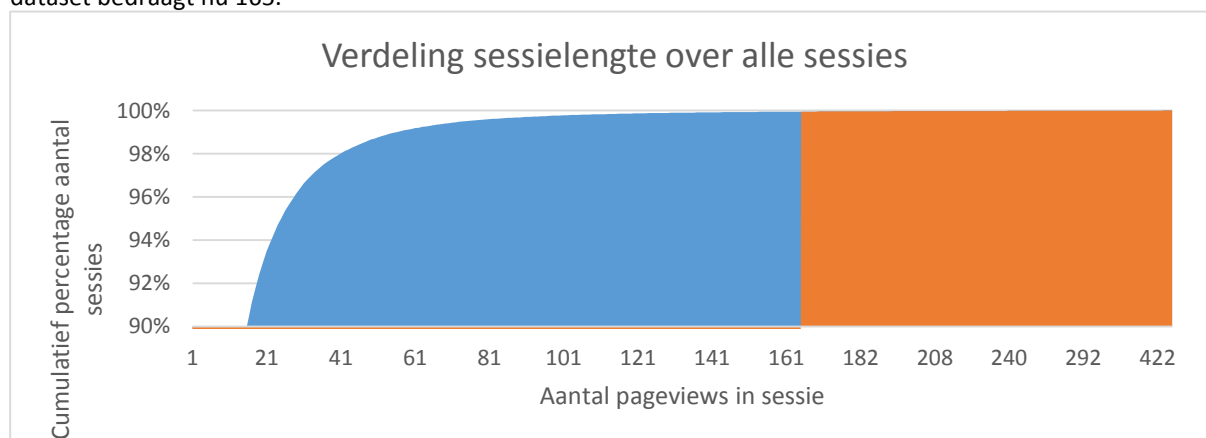
4.2.1.5 SESSIE-HERDEFINIËRING

Als een bezoeker meerdere malen op een dag de website van Bol.com bezoekt, wordt hij door de cookies geregistreerd als eenzelfde bezoeker. Hierdoor kan het voorkomen dat een gebruiker de website bezoekt, dan een aantal uren niet actief is, en vervolgens de website nogmaals bezoekt. Dit lijken op het eerste gezicht twee verschillende bezoeken, maar toch worden deze als één sessie geregistreerd.

Voor het analyseren van de data worden deze twee bezoeken uit elkaar getrokken. Hierbij geldt dat wanneer een gebruiker minimaal een uur inactief is geweest, het volgende uur als nieuwe sessies wordt geregistreerd. Op deze manier ontstaan er meerdere sessies dan oorspronkelijk. Het nieuwe aantal sessies wordt nu 445.391 ten opzichte van de eerdere 440.277.

4.2.2 OUTLIERS

Ook de bovenste 0,05% van de sessies qua sessielengte is uit de data verwijderd, omdat de betrouwbaarheid van deze sessies zeer laag zijn. Een korte analyse toont aan dat het lijkt alsof een groot aantal sessies door robots worden gemaakt. Dit is te zien aan de tijden waarop de pageviews worden bezocht (constante tussentijden) en aan de beperkte variatie in de pageviews. De 0,05% omvat in totaal 181 sessies, 101 verschillende sessielengtes en 42.163 pageviews, als ook te zien is in Figuur 4.2. De maximale sessielengte in de dataset bedraagt nu 165.



Figuur 4.2 - Outliers op basis van sessielengte. Het oranje vlak geeft aan welke sessies uit de dataset zijn verwijderd.

4.2.3 RESULTAAT

Door het filteren van de dataset is er een nieuwe dataset ontstaan. Tabel 4.5 geeft een samenvatting van de records van de nieuwe dataset.

Tabel 4.5 – Overzicht dataset voor en na de preparatie.

	Voor outlier-detectie	Na outlier-detectie
Totaal aantal sessies	357.907	357.726
Totaal aantal pageviews	2.401.472	2.359.309
Gemiddeld aantal pageviews per sessie	6,71	6,60

4.3 ANALYSE BOL-DATA

In deze paragraaf wordt een analyse van de geprepareerde dataset van Bol.com gemaakt. In de finale dataset zijn een aantal nieuwe variabelen toegevoegd. Dit zijn variabelen als conversie, sessielengte, volgorde van pageviews binnen een sessie en waarde die aangeeft of een pagina eerder is bezocht gedurende de sessie. Aan de hand van deze en de oorspronkelijk variabelen heeft de daadwerkelijke analyse van de Bol-data plaatsgevonden. Hierbij is het belangrijk op te merken dat wanneer er gesproken wordt over laadtijd het de tijd tussen de navigationStart en het loadEventEnd betreft.

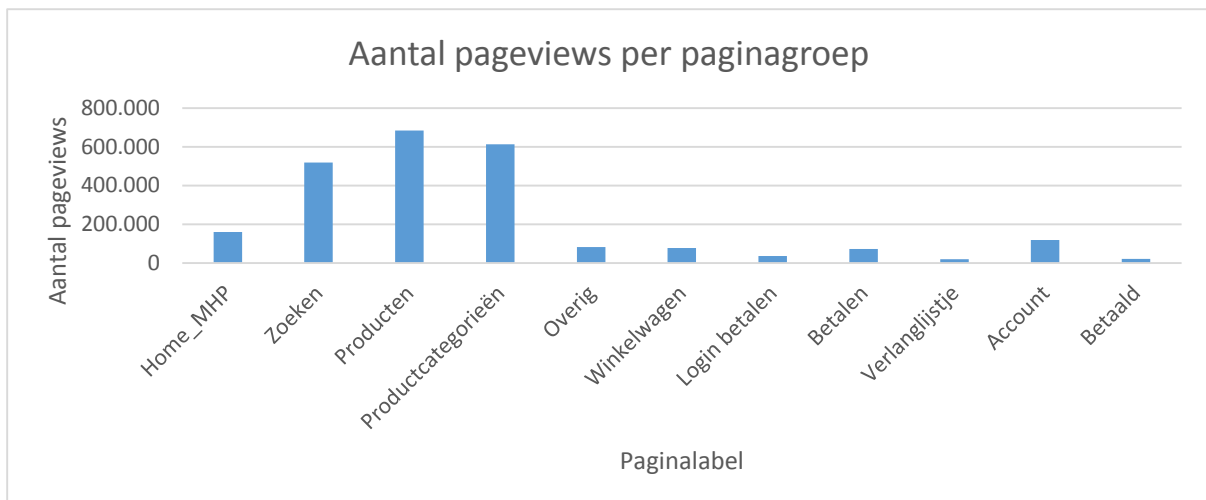
4.3.1 PAGINAGROEP

Om de verschillende stadia van een websitebezoek te kunnen categoriseren is er op basis van het door Bol toegekende RUM-label een nieuw label in de vorm van een paginagroep aan elke pageview toegevoegd. Deze paginagroepen zijn opgesteld op basis van de verschillende (navigatie)-stappen binnen het conversieproces, zie Tabel 4.6.

Elke paginagroep kent een eigen ID. Het gebruik van integers versnelt het zoekproces in de database ten opzichte van het zoeken op basis van strings. Ook is er per pageview vastgesteld tot welke paginagroep de vorige pagina behoorde en deze waarde is ook opgeslagen in de database. In de grafiek (Figuur 4.3) is de verdeling van de pageviews over de verschillende paginagroepen te zien.

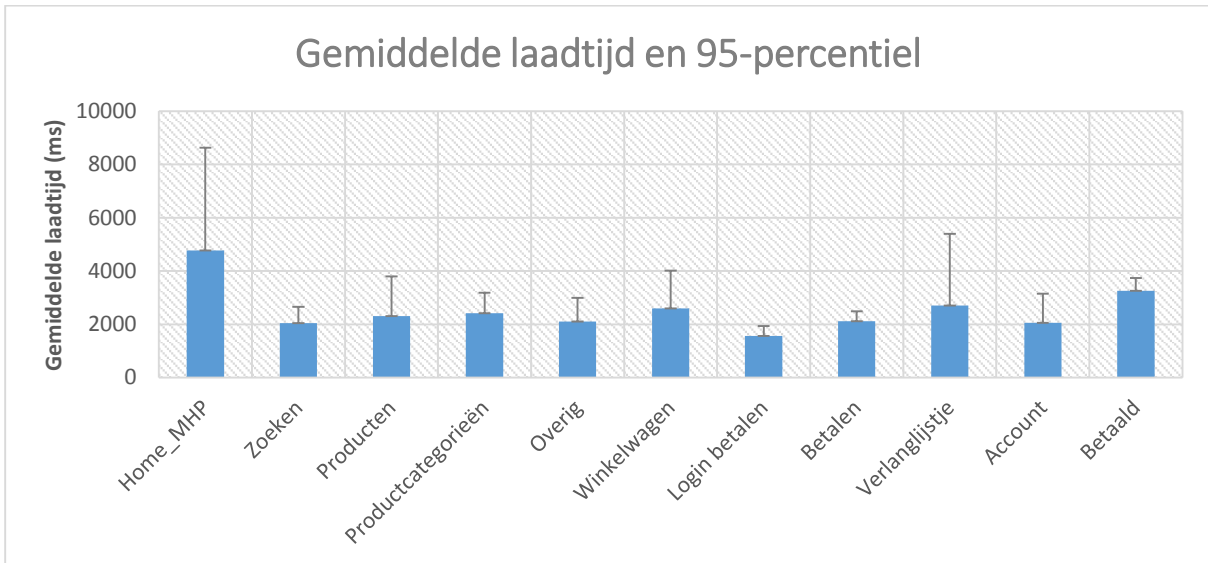
Tabel 4.6 - Overzicht paginagroepen.

Paginagroep	ID	Beschrijving	Aantal
Home_MHP	1	De "Main Home Page" van Bol.com	159.196
Zoeken	2	Overzicht van producten binnen zoekresultaten	518.436
Producten	3	Product-details-pagina's	682.903
Productcategorieën	4	Overzichten van producten binnen bepaalde categorie	612.884
Overig	5	Alle overige pagina's	82.272
Winkelwagen	6	Overzicht van producten in de winkelwagen	77.002
Login betalen	7	Het inloggen gedurende het betaalproces	35.894
Betalen	8	Pagina waarop klant gaat afrekenen	73.135
Verlanglijstje	9	Overzicht van producten op verlanglijstje van een klant	19.072
Account	10	Het klikken binnen de account-pagina's	118.985
Betaald	11	De pagina na de betaling van een bestelling	21.693



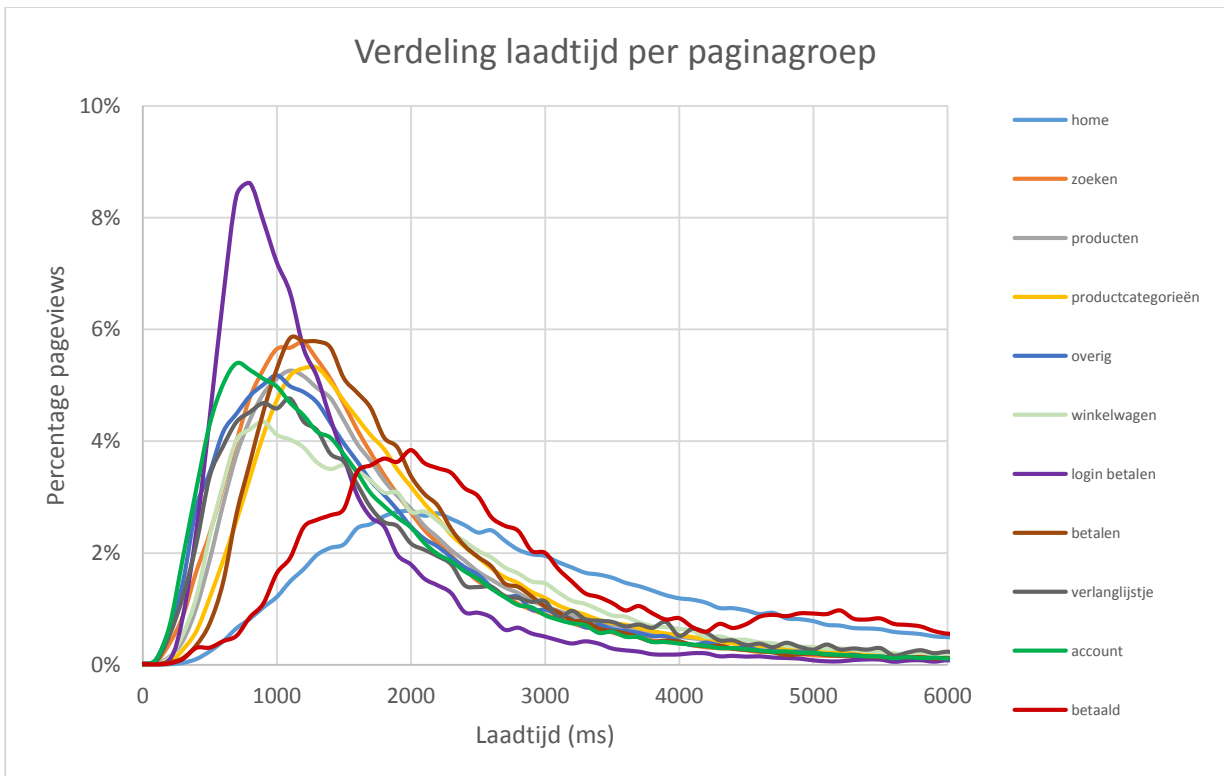
Figuur 4.3 – Overzicht aantal pageviews per paginagroep.

Figuur 4.4 geeft de gemiddelde laadtijd en het 95-percentiel van de pageviews per paginagroep weer. Hierbij valt op dat de homepagina de hoogste gemiddelde laadtijd alsmede de hoogste variatie in de laadtijd kent. De paginagroep die gemiddeld de laagste laadtijd heeft is de pagina waarop dient te worden ingelogd alvorens het betaalproces kan worden gestart.



Figuur 4.4 - Gemiddelde laadtijd per paginagroep.

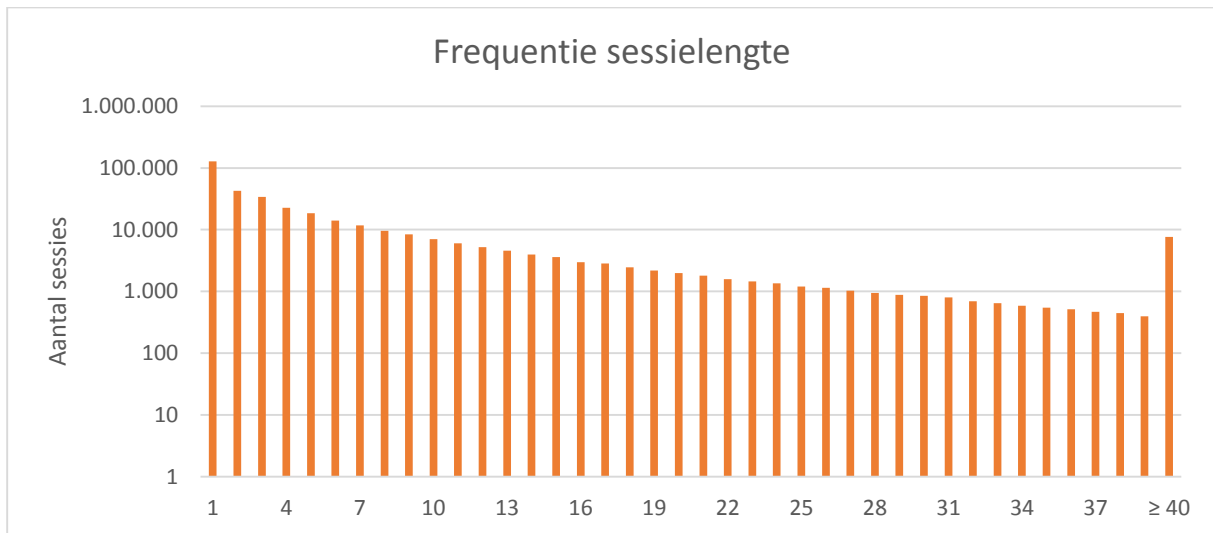
De onderstaande grafiek (Figuur 4.5) geeft de gemiddelde laadtijd per paginagroep weer evenals de variatie van de gemiddelde laadtijd per paginagroep. De meest stabiele paginagroep lijkt hier het inloggen voor aanvang van het betaalproces. De verdelingen van de laadtijd van de paginagroepen lijken redelijk overeen te komen. Hierbij worden de paginagroepen home, login-betalen en betaald uitgezonderd.



Figuur 4.5 - Verdeling laadtijd per paginagroep.

4.3.1.1 SESSIELENGTE

De lengte van de sessie is het aantal pageviews dat een sessie telt. Het volgende figuur (Figuur 4.6) geeft een overzicht van de frequentie waarop een zekere sessielengte voor komt. Hierbij is de verticale as logaritmisch verdeeld.

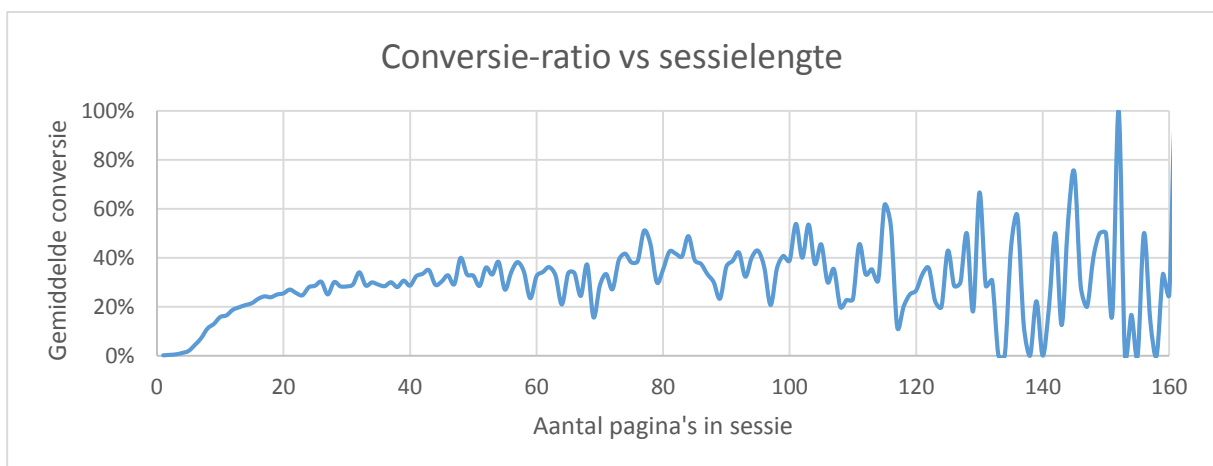


Figuur 4.6 – Overzicht van de frequentie van de verschillende sessielengtes.

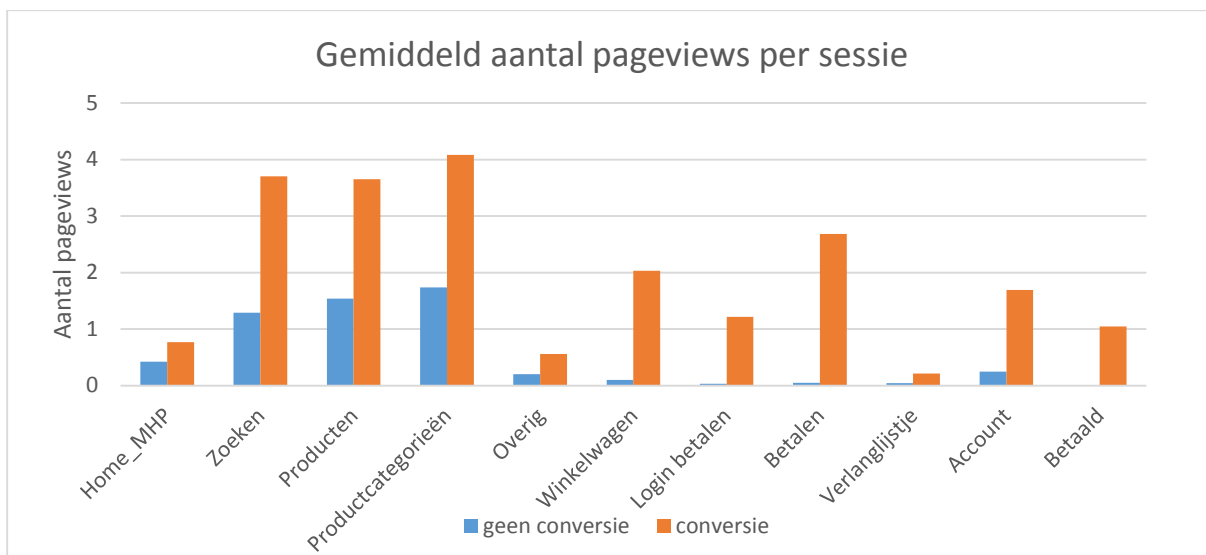
4.3.1.2 CONVERSIE-RATIO

De conversie-ratio wordt bepaald aan de hand van de paginagroep 'betaald'. Als een gebruiker een pagina uit deze paginagroep heeft bezocht, wordt er aangenomen dat er daadwerkelijk iets is afgenomen bij Bol.com. Een bezoeker kan namelijk alleen op deze pagina komen wanneer een bestelling volledig is afgerond. De paginagroep bevat slechts één pagina, namelijk de 'betaald'-pagina van Bol.com.

Het aantal sessies waarin tot conversie wordt overgegaan betreft 5.8% (20.644 van de 357.726 sessies). De grafieken hieronder pogen inzicht te geven in de verschillende beschikbare factoren die de conversie beïnvloeden. Zo bevat de gemiddelde sessie van een converterende bezoeker meer pagina's dan de sessie van een niet-converterende bezoeker (Figuur 4.7). Een van de mogelijke oorzaken is dat de converterende klant naast het vullen van zijn winkelmand ook het gehele betaalproces moet doorlopen. Ook is in Figuur 4.8 te zien dat een converterende bezoeker meer zoekt, meer productcategorieën bekijkt en vaker specifieke productgegevens ophaalt dan een niet-converterende bezoeker.



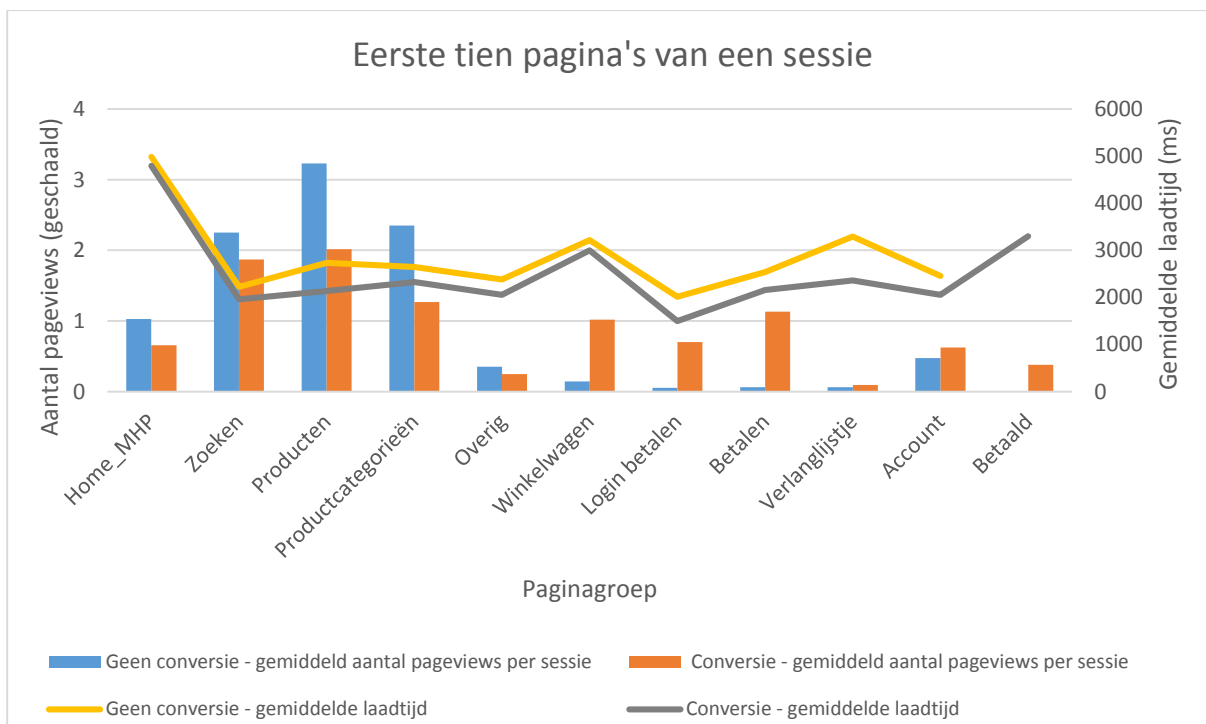
Figuur 4.7 – De conversie-ratio afhankelijk van de sessielengte.



Figuur 4.8 – Aantal pageviews per converterende en niet-converterende sessie.

Figuur 4.9 toont het gemiddeld aantal pageviews van de (maximaal) eerste tien pagina's van een sessie. Er wordt hierbij onderscheid gemaakt tussen de converterende en niet-converterende sessies. Omdat een groot aantal sessies een sessielengte kleiner dan tien pageviews heeft, is het gemiddeld aantal pageviews geschaald. Op deze manier is het mogelijk de converterende en niet-converterende sessies met elkaar te vergelijken.

Ook wordt de gemiddelde laadtijd per paginagroep weergegeven. Te zien is dat deze voor de converterende sessies voor iedere paginagroep lager ligt dan de gemiddelde laadtijd voor de niet-converterende sessies.



Figuur 4.9 – Pageview-gegevens van de maximaal eerste tien pagina's van een sessie onderverdeeld naar converterende en niet-converterende sessies.

4.3.2 VOLGORDE VAN PAGINA'S BINNEN EEN SESSIE

De volgorde van de pageviews binnen een sessie is gedefinieerd middels een nieuwe variabele. Ditzelfde geldt voor de pagina's die als eerste en als laatste worden bezocht tijdens het websitebezoek.

4.3.2.1 NAVIGATIE TUSSEN PAGINAGROEPEN

Er is onderscheid te maken op basis van de navigatie tussen webpagina's. Tabel 4.7 en 4.8 tonen de wijze waarop de bezoekers navigeren tussen paginagroepen. Hierbij geeft Tabel 4.7 het percentage navigaties van het totaal aantal navigaties in de dataset aan. Met een navigatie wordt een overgang van paginagroep X naar paginagroep Y bedoeld. Wanneer men start met de navigatie op een website, wordt dit aangegeven met de paginagroep 'start'. Indien men de website verlaat wordt dit aangegeven met de paginagroep 'eind'.

Uit de data blijkt dat niet converterende bezoekers de website het vaakst verlaten wanneer zij zich op een productpagina bevinden. Daarnaast maakt deze groep bezoekers vaker de overgang van een productcategorie naar een specifiek product dan de bezoekers die daadwerkelijk tot conversie overgaan.

Tabel 4.7 – Overzicht van kansen van paginagroep x naar paginagroep, gebaseerd op totaal aantal navigaties. De kleuren geven percentages groter of gelijk aan 5% aan.

Paginagroep van \ naar	Home_MHP	Zoeken	Producten	Productcategorieën	Overig	Winkelwagen	Login betalen	Betalen	Verlanglijstje	Account	Betaald	Eind
geen conversie												
Home_MHP	0%	3%	0%	2%	0%	0%	0%	0%	0%	0%	0%	1%
Zoeken	0%	9%	6%	1%	0%	0%	0%	0%	0%	0%	0%	2%
Producten	0%	4%	8%	4%	1%	0%	0%	0%	0%	0%	0%	7%
Productcategorieën	0%	2%	5%	13%	0%	0%	0%	0%	0%	0%	0%	3%
Overig	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	1%
Winkelwagen	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Login betalen	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Betalen	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Verlanglijstje	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Account	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%	0%	1%
Betaald	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Start	5%	1%	6%	2%	1%	0%	0%	0%	0%	0%	0%	0%
conversie												
Home_MHP	0%	2%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%
Zoeken	0%	7%	6%	1%	0%	1%	0%	0%	0%	0%	0%	0%
Producten	0%	4%	5%	3%	1%	3%	0%	0%	0%	0%	0%	0%
Productcategorieën	0%	1%	3%	10%	0%	1%	0%	0%	0%	0%	0%	0%
Overig	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Winkelwagen	0%	1%	1%	0%	0%	2%	4%	1%	0%	0%	0%	0%
Login betalen	0%	0%	0%	0%	0%	0%	1%	3%	0%	0%	0%	0%
Betalen	0%	0%	0%	0%	0%	0%	0%	7%	0%	0%	4%	0%
Verlanglijstje	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Account	0%	0%	0%	0%	0%	0%	0%	0%	0%	4%	0%	1%
Betaald	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	2%
Start	2%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%
verschil: conversie - geen conversie												
Home_MHP	0%	-1%	0%	-1%	0%	0%	0%	0%	0%	0%	0%	-1%
Zoeken	0%	-2%	0%	0%	0%	1%	0%	0%	0%	0%	0%	-2%
Producten	0%	0%	-2%	-1%	0%	3%	0%	0%	0%	0%	0%	-7%
Productcategorieën	0%	0%	-1%	-4%	0%	1%	0%	0%	0%	0%	0%	-2%
Overig	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	-1%
Winkelwagen	0%	0%	1%	0%	0%	1%	3%	1%	0%	0%	0%	0%
Login betalen	0%	0%	0%	0%	0%	0%	1%	3%	0%	0%	0%	0%
Betalen	0%	0%	0%	0%	0%	0%	0%	6%	0%	0%	4%	0%
Verlanglijstje	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Account	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	1%
Betaald	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	2%
Start	-3%	-1%	-5%	-2%	0%	0%	0%	0%	0%	0%	0%	0%

Tabel 4.8 toont het percentage navigaties tussen twee paginagroepen op basis van het totaal aantal navigaties van een zekere paginagroep. Dit betekent dat het rijtotaal 100% bedraagt. Op deze manier wordt inzichtelijk wat de meest gekozen paginagroepen zijn gegeven de huidige paginagroep waarop een bezoeker zich bevindt.

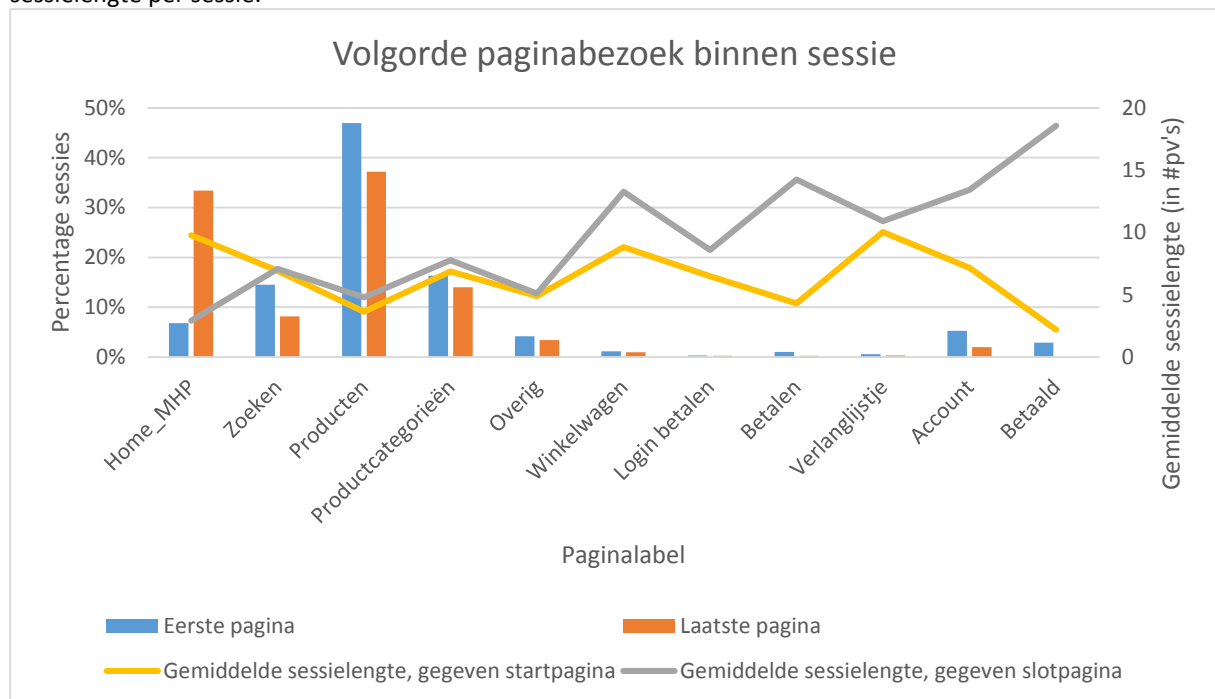
Te zien is dat het vaak voorkomt dat twee dezelfde paginagroepen direct na elkaar worden bezocht. Dit kan te wijten zijn aan het verversen van een pagina, maar kan ook worden veroorzaakt doordat klanten bijvoorbeeld van een productpagina direct naar een andere productpagina navigeren. Ook valt op dat de converterende bezoekers vaker via de homepagina binnenkomen dan de niet-converterende bezoekers en dat deze laatste groep ook vaker achter elkaar het winkelmandje bezoeken. Een logisch gevolg van het onderscheid op basis van conversie is dat de slotpagina van deze groepen zeer van elkaar verschilt.

Tabel 4.8 – Overzicht van kansen van paginagroep x naar paginagroep, gebaseerd op een rij-totaal van 100%. De kleuren geven percentages groter of gelijk aan ±10% aan.

Paginagroep van \ naar	Home_MHP	Zoeken	Producten	Productcategorieën	Overig	Winkelwagen	Login betalen	Betalen	Verlanglijstje	Account	Betaald	Eind
geen conversie												
Home_MHP	3%	40%	5%	24%	4%	1%	0%	0%	1%	5%	0%	17%
Zoeken	2%	47%	31%	6%	1%	1%	0%	0%	0%	0%	0%	12%
Producten	2%	16%	30%	17%	4%	2%	0%	0%	1%	1%	0%	28%
Productcategorieën	1%	7%	21%	57%	2%	0%	0%	0%	0%	1%	0%	17%
Overig	2%	12%	29%	16%	14%	3%	0%	0%	0%	2%	0%	21%
Winkelwagen	2%	7%	12%	7%	2%	30%	18%	6%	2%	2%	0%	12%
Login betalen	1%	1%	1%	2%	0%	9%	31%	40%	0%	6%	0%	9%
Betalen	3%	1%	1%	2%	1%	6%	1%	58%	0%	8%	0%	21%
Verlanglijstje	3%	10%	24%	7%	2%	6%	0%	0%	29%	7%	0%	14%
Account	2%	2%	4%	4%	2%	1%	0%	0%	1%	68%	0%	15%
Betaald	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Start	32%	8%	38%	14%	4%	1%	0%	0%	0%	2%	0%	0%
conversie												
Home_MHP	3%	47%	7%	23%	3%	6%	0%	0%	2%	9%	0%	2%
Zoeken	1%	46%	36%	6%	2%	8%	0%	0%	0%	0%	0%	1%
Producten	2%	22%	30%	16%	6%	17%	0%	0%	1%	2%	0%	2%
Productcategorieën	1%	9%	20%	60%	2%	4%	0%	0%	0%	1%	0%	1%
Overig	1%	15%	31%	17%	11%	19%	0%	0%	1%	3%	0%	2%
Winkelwagen	1%	6%	8%	5%	1%	21%	41%	14%	2%	2%	0%	0%
Login betalen	0%	0%	0%	0%	0%	5%	27%	62%	0%	5%	0%	0%
Betalen	1%	0%	0%	1%	0%	2%	1%	58%	0%	1%	37%	0%
Verlanglijstje	1%	7%	23%	5%	2%	32%	0%	0%	21%	6%	0%	2%
Account	2%	4%	5%	5%	1%	4%	0%	3%	1%	56%	1%	18%
Betaald	2%	2%	4%	3%	1%	2%	0%	0%	1%	36%	1%	48%
Start	49%	8%	18%	9%	2%	5%	2%	2%	0%	3%	1%	0%
Verskil: conversie – geen conversie												
Home_MHP	-1%	7%	2%	-1%	-1%	5%	0%	0%	1%	4%	0%	-15%
Zoeken	-1%	-1%	5%	0%	0%	7%	0%	0%	0%	0%	0%	-11%
Producten	0%	6%	1%	0%	2%	16%	0%	0%	1%	1%	0%	-26%
Productcategorieën	0%	2%	0%	3%	0%	3%	0%	0%	0%	1%	0%	-10%
Overig	-1%	2%	2%	0%	-3%	17%	0%	0%	0%	1%	0%	-20%
Winkelwagen	-1%	-1%	-4%	-2%	-1%	-10%	23%	8%	0%	-1%	0%	-11%
Login betalen	0%	-1%	0%	-2%	0%	-4%	-4%	22%	0%	-1%	0%	-9%
Betalen	-2%	0%	-1%	-1%	0%	-4%	-1%	0%	0%	-7%	37%	-21%
Verlanglijstje	-2%	-2%	-1%	-2%	0%	27%	0%	0%	-7%	-1%	0%	-12%
Account	0%	1%	2%	1%	0%	3%	0%	2%	0%	-12%	1%	3%
Betaald	2%	2%	4%	3%	1%	2%	0%	0%	1%	36%	1%	48%
Start	17%	0%	-20%	-5%	-2%	5%	2%	2%	0%	1%	1%	0%

4.3.2.2 LANDINGS- EN SLOT PAGINA

Figuur 4.9 geeft een beeld van de verdeling van deze eerste (landings-) en laatste pagina per paginagroep. Hierbij geven de kolommen de verhouding van het aantal pageviews aan en de lijnen de gemiddelde sessielengte per sessie.



Figuur 4.9 - Overzicht van de volgorde van pageviews binnen een sessie.

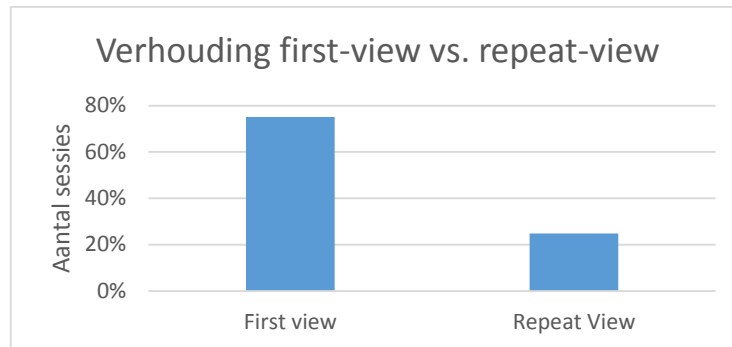
Opvallend in deze grafiek is dat de gemiddelde sessielengte het hoogst is wanneer de landingspagina de homepagina of het verlanglijstje is. Met de kanttekening dat er zeer weinig sessies zijn die beginnen op de pagina met het verlanglijstje.

Wanneer er op een specifieke productpagina gestart wordt zal de gemiddelde sessielengte het kortst zijn, met uitzondering van de pagina waarop een bestelling is afgerond. Het lijkt er echter op dat wanneer deze pagina als landingspagina wordt bezocht iets anders aan de hand is. Het is mogelijk dat deze pagina in een andere browser wordt geopend, waardoor het lijkt alsof er een nieuwe sessie wordt gestart. Het gaat hierbij slechts om 1% van alle pageviews van de 'betaald'-pagina's.

De slotpagina van een sessie is in de meeste gevallen de homepagina of een specifieke productpagina. Bij beide situaties geldt dat de gemiddelde sessielengte relatief laag is. De gemiddelde sessielengte is het hoogst wanneer men eindigt op de pagina waarop de bestelling is afgerond. Dit komt overeen met de analyse dat een converterende klant (en dus een klant die de bestelling afrond) gemiddeld meer pagina's bezoekt dan een niet-converterende klant.

4.3.2.3 REPEAT VIEW

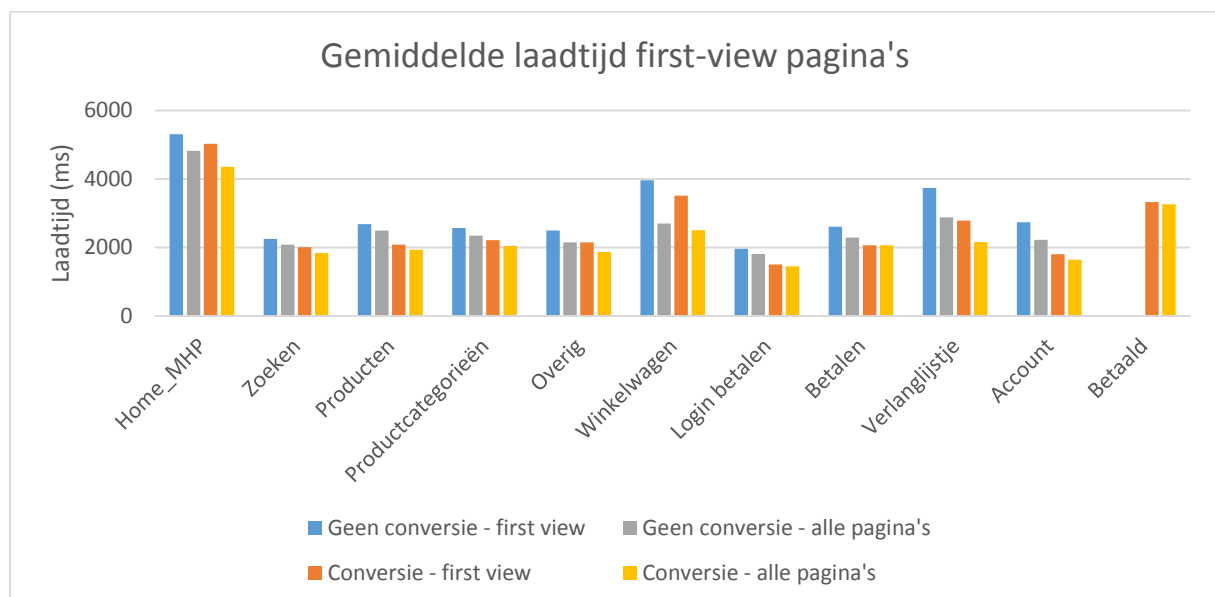
De variabele repeat-view is ook toegevoegd aan de dataset. Deze waarde geeft aan of een URL al dan niet eerder in de sessie is bezocht. Als een pagina voor de eerste maal wordt bezocht spreken we van een first-view, anders van een repeat-view. Een overzicht van de verhouding van first-view-pageviews en repeat-view-pageviews is te vinden in Figuur 4.10.



Figuur 4.10 - Overzicht verhouding first- en repeat-view.

Het onderscheid op basis van first-view en repeat-view is van belang omdat de laadtijd van repeat-view-pagina's lager ligt dan de laadtijd voor diezelfde pagina's die voor de eerste maal worden bezocht (Figuur 4.12). Dit wordt veroorzaakt doordat er bij een repeat-view meer objecten gecached kunnen worden waardoor er minder data te hoeft worden opgehaald bij de webserver. Ook wordt in Figuur 4.11 onderscheid gemaakt tussen de laadtijd van first-view en repeat-view van converterende en niet-converterende sessies.

Wanneer er gekeken wordt naar het verschil in laadtijd tussen de pagina's in een converterende sessie en de pagina's in een niet-converterende sessie, is waarneembaar dat de laadtijd van de eerste categorie sessies altijd lager ligt dan de laadtijd van de tweede categorie sessies.



Figuur 4.10 – Overzicht gemiddelde laadtijd first-view pagina's met onderscheid tussen converterende en niet-converterende sessies.

4.4 CONCLUSIE DATA-ANALYSE

Op basis van de bovenstaande data-analyse lijkt het van belang de focus te leggen op het onderscheid tussen converterende en niet-converterende bezoekers. Uit de data-analyse blijkt dat factoren als laadtijd, sessielengte en pagina-navigatie een rol spelen bij de kans dat een klant al dan niet tot conversie over gaat. Deze factoren vormen dan ook de basis van de methodiek die in het volgende hoofdstuk (Hoofdstuk 5) beschreven wordt.

5 METHODIEK

In dit hoofdstuk wordt nader ingegaan op de methodiek om bij achterblijvende performance de conversie te verhogen. In eerste instantie zal deze methodiek uit twee onderdelen bestaan. Het eerste onderdeel is gebaseerd op de navigatie door de website. Het tweede onderdeel is gebaseerd op de door de MW-tag gemeten laadtijden.

Het eerste onderdeel is van belang om op deze manier de potentieel converterende klanten te onderscheiden van de klanten die überhaupt niet van plan zijn aankopen te doen. Door dit onderscheid te maken is het mogelijk een bepaalde boodschap alleen naar klanten te sturen die potentie hebben om tot conversie over te gaan.

Het tweede onderdeel dient te bepalen wanneer er een boodschap dient te worden gestuurd. Wanneer is de performance zo laag dat men afhaakt en zal men door berichtgeving een grotere kans hebben toch op de website te blijven.

5.1 ORC₁-WAARDE: NAVIGATIE DOOR DE WEBSITE

De informatie die wordt vergaard met betrekking tot de navigatie door een gebruiker tussen verschillende paginagroepen kan worden gebruikt om bezoekers te onderscheiden. Wanneer bezoekers de door converterende-klanten meest gekozen weg inslaan, lijkt de kans groter dat zij ook gaan converteren. Zo blijkt dat de kans op conversie groter is wanneer iemand als eerste de homepagina bezoekt dan iemand die als eerste een productpagina bezoekt. Dit gegeven wordt gebruikt om de ORC₁-waarde te berekenen. ORC staat hier voor Online Representative Customer en deze waarde wordt nader toegelicht in Paragraaf 5.1.3.

Om de meeste gangbare route voor een converterende klant te berekenen, wordt er gebruik gemaakt van voorwaardelijke kansen. Wanneer er in nader te bepalen mate wordt afgeweken van dit door-converterende-klanten-meest-gekozen pad, lijkt het er op dat de kans op conversie afneemt. In deze paragraaf wordt bekeken hoe het ideale pad van een bezoeker er uit ziet en wat er dient te gebeuren wanneer er door een bezoeker in een zekere mate van dat pad wordt afgeweken.

5.1.1 VOORWAARDELIJKE KANS

Een voorwaardelijk kans is de kans dat een gebeurtenis plaatsvindt gegeven dat er een andere gebeurtenis plaatsvindt. De kans dat een gebeurtenis C plaatsvindt, gegeven dat een andere gebeurtenis A plaatsvindt wordt gedefinieerd in Vergelijking 5.1.

$P(C A) = \frac{P(AC)}{P(A)}$	
met	C en A – situatie of object
	<i>F</i> 's - de eigenschapsvariabelen

Vergelijking 5.1 – Definitie van een voorwaardelijke kans (Olshausen 2004).

Om het meest gangbare pad van een converterende bezoeker te bepalen, wordt de mogelijkheid dat een bezoeker van pagina X naar pagina Y navigeert bepaald aan de hand van een voorwaardelijke kans. De definitie van deze overgangskans wordt weergegeven in Vergelijking 5.2. Het toepassen van deze methode op de huidige dataset resulteert in de data weergegeven in Tabel 4.8.

$$p_{xy} = P(PG_{v+1} = y | PG_v = x) = \frac{z_{xy}}{\sum_{j=1}^k z_{xj}}$$

met

- PG - de paginagroep waar een bezoeker zich in bevindt. Waarbij v de huidige pageviews representeert en v+1 de volgende pageview representeert
- z_{xy} - het aantal pageviews van het totale aantal pageviews uit de dataset waarbij van paginagroep x naar paginagroep y wordt genavigeerd
- k - het totaal aantal gedefinieerde paginagroepen

Vergelijking 5.2 – Voorwaardelijke kans: kans dat een bezoeker naar paginagroep y te navigeert, gegeven dat de bezoeker zich nu op paginagroep x bevindt.

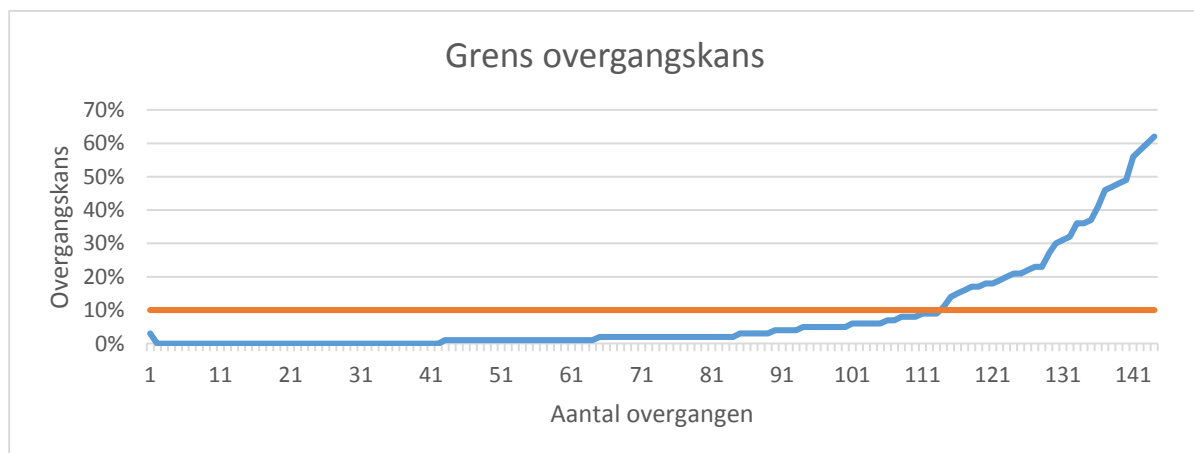
5.1.2 HET BEPALEN VAN DE OVERGANGSKANSEN

Iedere pageview krijgt vervolgens een bijbehorende overgangskans toegewezen. Dit is echter niet de overgangskans die hierboven beschreven is en weergegeven in Tabel 4.8, maar een bewerking op deze overgangskans. De reden hiervoor wordt toegelicht in Subparagraaf 5.1.3.

Door onderscheid te maken tussen ‘normale’ en ‘afwijkende’ overgangen kan er onderscheid gemaakt worden tussen sessies. Een q-overgangskans wordt gedefinieerd door Vergelijking 5.3. De grens is bij 0,1 gesteld om zo een initiële waarde te creëren die de waarschijnlijk- van de minder-waarschijnlijk genomen overgangen van elkaar onderscheid. De keuze voor 0,1 is gebaseerd op het feit dat er rond deze waarde een tweedeling van de dataset lijkt te ontstaan (Figuur 5.1).

$$q_{xy} = \begin{cases} \sum_{j \in S} p_{xj} & S = \{j: p_{xj} \geq 0,1\} & \text{als } p_{xy} \geq 0,1 \\ \sum_{j \in S} p_{xj} & S = \{j: p_{xj} < 0,1\} & \text{als } p_{xy} < 0,1 \end{cases}$$

Vergelijking 5.3 – Definitie van de q-overgangskans.



Figuur 5.1 – Bepalen van de grens voor het berekenen van de q-overgangskans. Bij 10% lijkt er een tweedeling van de dataset te ontstaan.

Op deze manier hebben de veel voorkomende overgangen een grotere waarde dan de overgangen die weinig tot zelden voorkomen. De q-overgangskansen van de converterende-pageviews worden in Tabel 5.1 weergegeven.

Tabel 5.1 – Overzicht q-overgangskansen op basis van de converterende sessies.

Paginagroep van \ naar	Home_MHP	Zoeken	Producten	Productcategorieën	Overig	Winkelwagen	Login betalen	Betalen	Verlanglijstje	Account	Betaald	Eind
Home_MHP	31%	69%	31%	69%	31%	31%	31%	31%	31%	31%	31%	31%
Zoeken	18%	82%	82%	18%	18%	18%	18%	18%	18%	18%	18%	18%
Producten	13%	87%	87%	87%	13%	87%	13%	13%	13%	13%	13%	13%
Productcategorieën	20%	20%	80%	80%	20%	20%	20%	20%	20%	20%	20%	20%
Overig	7%	93%	93%	93%	93%	93%	7%	7%	7%	7%	7%	7%
Winkelwagen	25%	25%	25%	25%	25%	75%	75%	75%	25%	25%	25%	25%
Login betalen	11%	11%	11%	11%	11%	11%	89%	89%	11%	11%	11%	11%
Betalen	5%	5%	5%	5%	5%	5%	5%	95%	5%	5%	95%	5%
Verlanglijstje	24%	24%	76%	24%	24%	76%	24%	24%	76%	24%	24%	24%
Account	26%	26%	26%	26%	26%	26%	26%	26%	26%	74%	26%	74%
Betaald	16%	16%	16%	16%	16%	16%	16%	16%	16%	84%	16%	84%
Start	67%	33%	67%	33%	33%	33%	33%	33%	33%	33%	33%	33%

Elke pageview uit de dataset krijgt de bijbehorende q-overgangskans toegewezen op basis van de huidige en de volgende pageview. Als het de eerste pageview van een sessie betreft wordt de q-overgangskans met als 'paginagroep van' 'start' gebruikt. De q-overgangskans van elk mogelijke paginagroep naar 'eind' zal nooit worden toegewezen, omdat op dat moment de klant de website heeft verlaten waardoor hij niet meer gemonitord kan worden.

5.1.3 ORC₁-WAARDE BEPALEN

De q-overgangskansen zijn dus hoger voor de overgangen die vaak ($\geq 10\%$) door converterende bezoekers worden gemaakt dan overgangen die veel minder vaak voorkomen ($< 10\%$). Door het gemiddelde van deze waarden gedurende de sessie bij te houden, is het mogelijk inzicht te krijgen in het type navigatie van de bezoeker. Wanneer de gemiddelde waarde zodanig klein wordt betekent dit dat het door de bezoeker genomen pad steeds onwaarschijnlijker leidt tot conversie. Door een ondergrens in te stellen voor dit gemiddelde kunnen de bezoekers in twee groepen worden opgedeeld waarbij de ene groep een kleinere kans op conversie lijkt te hebben dan de andere groep.

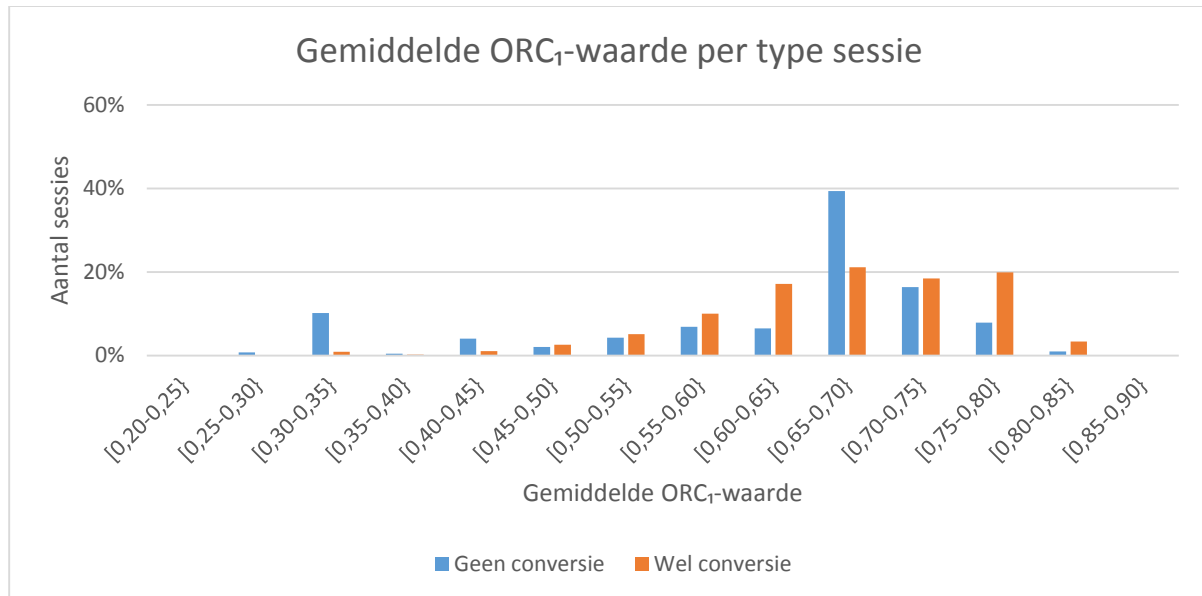
Voor deze gemiddelde waarde wordt de zogenaamde ORC₁-waarde geïntroduceerd. De index één geeft aan dat het de eerste toepassing is van de ORC-waarde. De ORC₁-waarde is het gemiddelde van alle q-waardes tot en met de huidige pageview binnen een sessie (Vergelijking 5.4).

$$ORC_1(o_{v,s}) = \frac{1}{o_{v,s}} \sum_{n=1}^{o_{v,s}} q^{s,n}$$

met $o_{v,s} \in \mathbb{N}$ $o_{v,s} =$ order van pageview v binnen een sessie s
 $q^{s,n} \in [0,1]$ $q^{s,n} =$ q-overgangskans behorend bij n^e pageview van sessie s

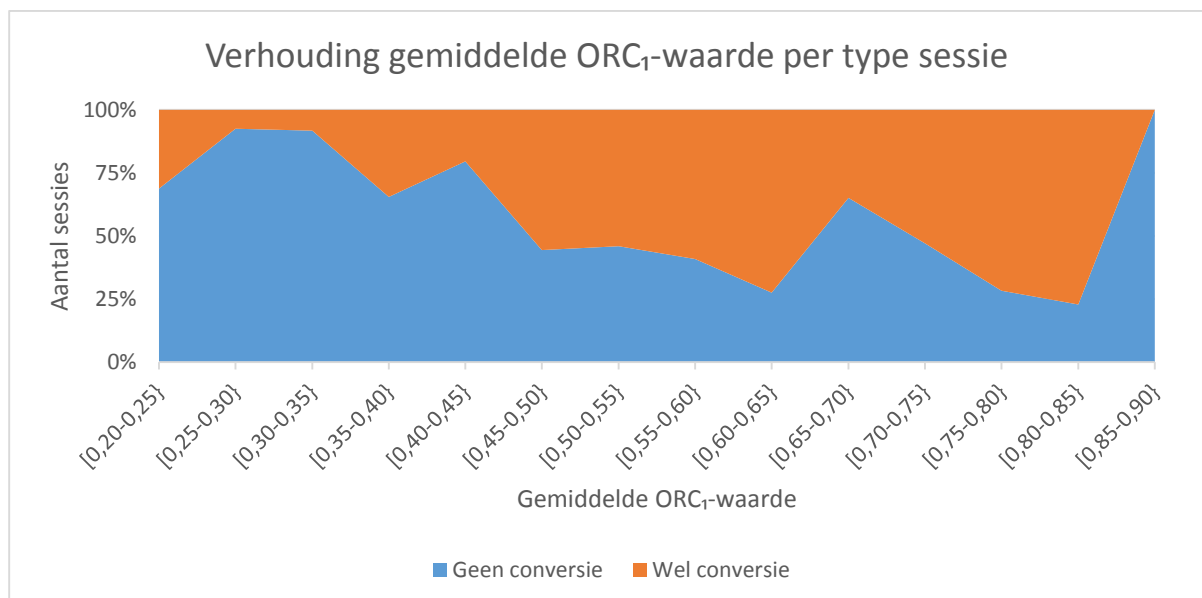
Vergelijking 5.4 – Definitie ORC₁-waarde.

Figuur 5.2 toont de gemiddelde ORC₁-waarde per type sessie. Hierbij is het aantal sessies relatief ten opzichte van het totaal aantal sessies waarbij onderscheid wordt gemaakt tussen converterende en niet-converterende sessies. De grafiek laat zien dat er relatief meer hoge ORC₁-waardes waar te nemen zijn voor de converterende-sessies, dan voor de niet-converterende sessies. Met name door de piek in het interval [0,30-0,35} voor de niet-converterende sessies is dit opvallend.



Figuur 5.2 – Overzicht van de gemiddelde ORC₁-waarde met onderscheid tussen converterende en niet-converterende sessies.

De tweede grafiek (Figuur 5.3) laat de verhouding van de gemiddelde ORC₁-waardes tussen converterende en niet-converterende bezoekers zien. Hoe hoger ORC₁-waardes zijn, des te groter de kans dat een sessie een converterende sessie betreft. Dit beeld onderschrijft dat de ORC₁-waarde voor de converterende sessies zich in de hogere regionen bevinden dan de ORC₁-waarde voor niet-converterende sessies. Het oppervlak van de grafiek is daar namelijk veel groter dan voor de lagere ORC₁-waardes.

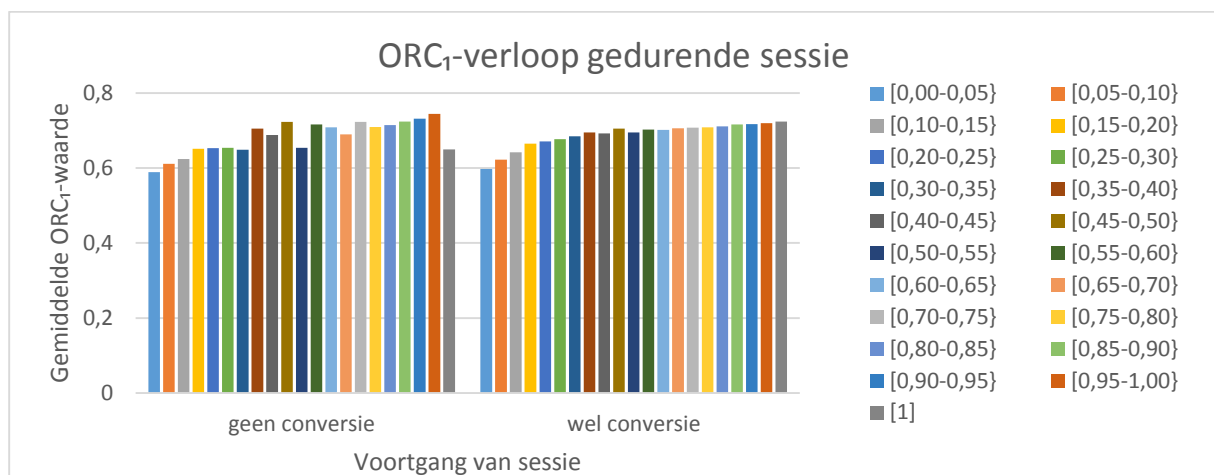


Figuur 5.3 – Overzicht van de verhouding van de gemiddelde ORC₁-waarde met onderscheid tussen converterende en niet-converterende sessies.

Figuur 5.4 geeft een beeld van het ORC_1 -verloop gedurende een sessie. In de grafiek is te zien dat naarmate een bezoeker zich in een verder stadium van zijn sessie bevindt, de gemiddelde ORC_1 -waarde toeneemt. Voor de converterende klanten is hier een duidelijker lijn in te zien dan voor de niet-converterende klanten.

Bij deze laatste groep zijn de twee meest negatieve uitschieters te zien bij het ORC_1 -interval $[0,50-0,55]$ en $[1]$. Deze uitschieters worden veroorzaakt door het grote aantal sessies van korte lengte: één of twee pageviews per sessie. De voortgang van de sessie bestaat dan uit een waarde van $1/1 = 1$, $1/2 = 0,5$ of $2/2 = 1$.

Een andere waarneming is dat het verloop van de ORC_1 -waarde stabiliseert naarmate de sessie vordert. Dit wordt verklaard doordat het gemiddelde gedurende de voortgang van de sessie, over steeds meer waardes wordt genomen. Een nieuwe q-overgangskans heeft dan steeds minder impact op de gemiddelde ORC_1 -waarde.



Figuur 5.4 – Overzicht van het ORC_1 -verloop gedurende een sessie met onderscheid tussen converterende en niet-converterende sessies.

5.1.4 HET BEPALEN VAN DE ONDERGRENS VAN DE ORC_1 -WAARDE

De ORC_1 -waarde kan worden ingezet om de converterende klanten van de niet-converterende klanten te onderscheiden. De resultaten in de vorige paragraaf laten zien dat de converterende sessies een hogere ORC_1 -waarde hebben dan de niet-converterende sessies. Dit gegeven kan worden gebruikt om bezoekers in één van de twee groepen in te delen. Hierbij heeft de groep met een ORC_1 -waarde boven de ondergrens een grotere kans op conversie, dan de groep met een ORC_1 -waarde onder de ondergrens. Op deze manier is het mogelijk alleen de klanten te benaderen die de meeste kans hebben om aankopen te doen.

De vraag is nu welke ondergrens er gekozen dient te worden zodat de kans op een verkeerde indicatie wordt geminimaliseerd. Hierbij ligt de nadruk op het minimaliseren van de kans op een 'false negative'. Dit houdt in dat een klant niet als potentieel converterend wordt gezien, terwijl dit wel het geval is. Een dergelijke error wordt ook wel een type I error genoemd. De kans op een type II error is de kans dat een bezoeker die niet over gaat tot conversie, alsnog als potentiële klant wordt gezien. Dit type error wordt ook wel een 'false positive' genoemd.

Uit de eerder genoemde resultaten blijkt dat wanneer de ondergrens van de ORC_1 -waarde bij 0,45 ligt, er 98% van de converterende klanten wordt bereikt. Wanneer deze ondergrens wordt gehanteerd, wordt ook 82% van de niet-converterende in deze groep ingedeeld. Als de grens hoger zou liggen, neemt het percentage converterende klanten dusdanig af dat er een grote kans is op false negatives. Daarentegen levert het weinig op wanneer de ondergrens naar beneden zal worden bijgesteld: het aantal converterende klanten neemt zeer weinig toe (van 98% naar 99%) terwijl het aantal false positives stijgt van 84% naar 89%.

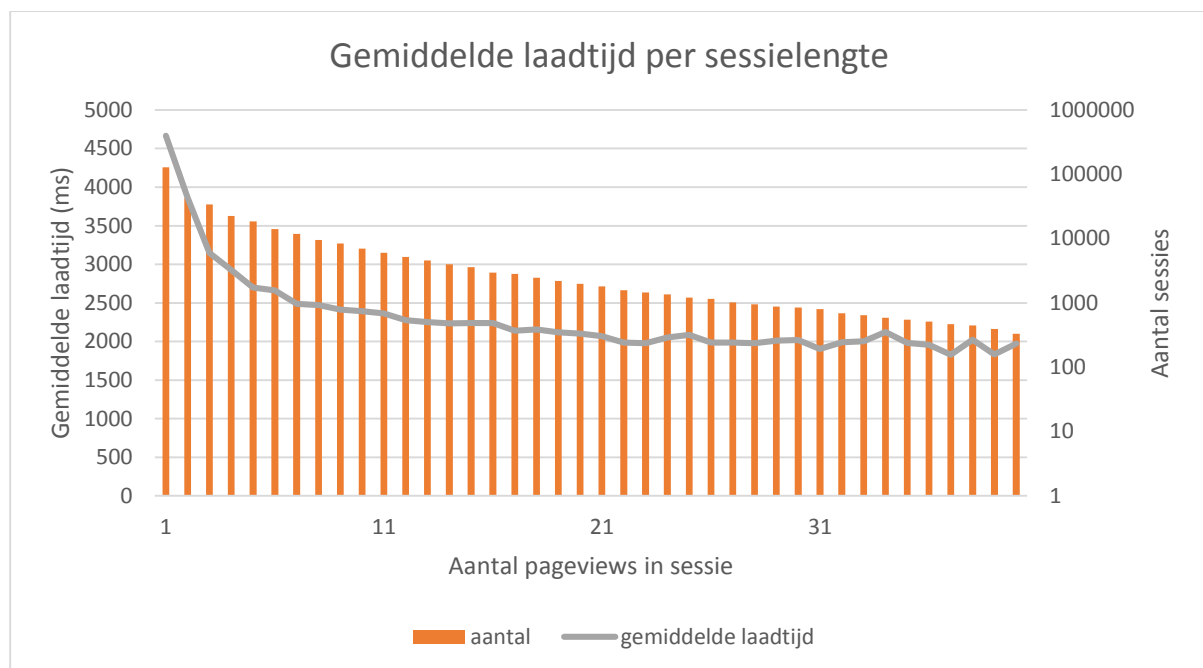
5.2 ORC₂-WAARDE: DE GEMIDDELDE LAADTIJD VERSUS DE SESSIELENGTE

In deze paragraaf wordt een tweede metriek voor de methodiek geïntroduceerd, de ORC₂-waarde. Deze waarde moet kunnen aangeven in welke gevallen de laadtijd zo hoog is, dat er ingegrepen dient te worden. Er zijn verschillende analyses verricht om te onderzoeken wat de beste definitie is voor deze waarde.

Er is onder andere gekeken naar de gemiddelde laadtijd over een zekere periode. Wanneer de huidige laadtijd teveel afwijkt van het gemiddelde, zou dit kunnen betekenen dat er iets aan de hand zou kunnen zijn. Echter wordt er op deze wijze geen rekening gehouden met de (re)acties van de bezoeker. Van te voren is niet aan te geven dat wanneer er een performance-degradatie plaatsvindt, bezoekers ook daadwerkelijk afhaken.

5.2.1 GEMIDDELDE LAADTIJD VERSUS SESSIELENGTE

De ORC₂-waarde is gebaseerd op de gemiddelde laadtijd versus de lengte van een sessie. Wanneer deze twee factoren tegenover elkaar worden weergegeven, lijkt er een exponentieel verband te bestaan (Figuur 5.5). De gemiddelde laadtijd neemt af naarmate de sessielengte toeneemt.



Figuur 5.5 - Overzicht gemiddelde laadtijd per sessielengte. De secundaire as geeft het aantal sessies weer.

In de eerste instantie is er een aantal verklaringen voor het verband te bedenken. Een eerste is dat de gemiddelde laadtijd lager wordt naarmate een sessie vordert omdat de cache wordt gebruikt bij het laden van de pagina. Dit betekent dat er minder objecten te hoeven worden opgehaald bij de server waardoor de laadtijd afneemt. Als er echter alleen gekeken wordt naar pagina's die al een keer in de sessie bezocht zijn (repeat-views) ontstaat hetzelfde verband.

Een tweede verklaring is dat pagina's die worden bezocht naarmate de sessie langer wordt (winkelmandje, betaalpagina, etc.) gemiddeld een hogere laadtijd hebben dan de pagina's die eerder voorkomen in de conversie-proces. Dit is onderzocht door alleen de pagina's behorend bij de eerste vier paginagroepen (home, zoeken, producten, productcategorieën) te bekijken. Ook hier komt eenzelfde exponentieel verband naar voren.

Een derde mogelijke verklaring is dat de laadtijd aan het begin van een sessie van belang is bij het besluiten op een website te blijven. Wanneer de laadtijd in het begin al te hoog wordt bevonden, is de kans dat men weg klikt vele malen groter.

5.2.2 EXPONENTIEEL VERBAND GEMIDDELDE LAADTIJD EN SESSIELENGTE

Het exponentiële verband dat lijkt te bestaan tussen de sessielengte en de gemiddelde laadtijd is nader onderzocht middels de MATLAB Curve Fitting Tool. De tool maakt gebruik van de 'Nonlinear Least Squares'-methode. De fit per paginagroep wordt weergegeven in Figuur 5.6. Hierbij is de oranje lijn de gefitte curve en representeert de blauwe lijn de werkelijke data. De fit is gebaseerd op Vergelijking 5.5.

$$y(x, a, b, c) = a * e^{-bx} + c$$

Met y - de gemiddelde laadtijd in ms

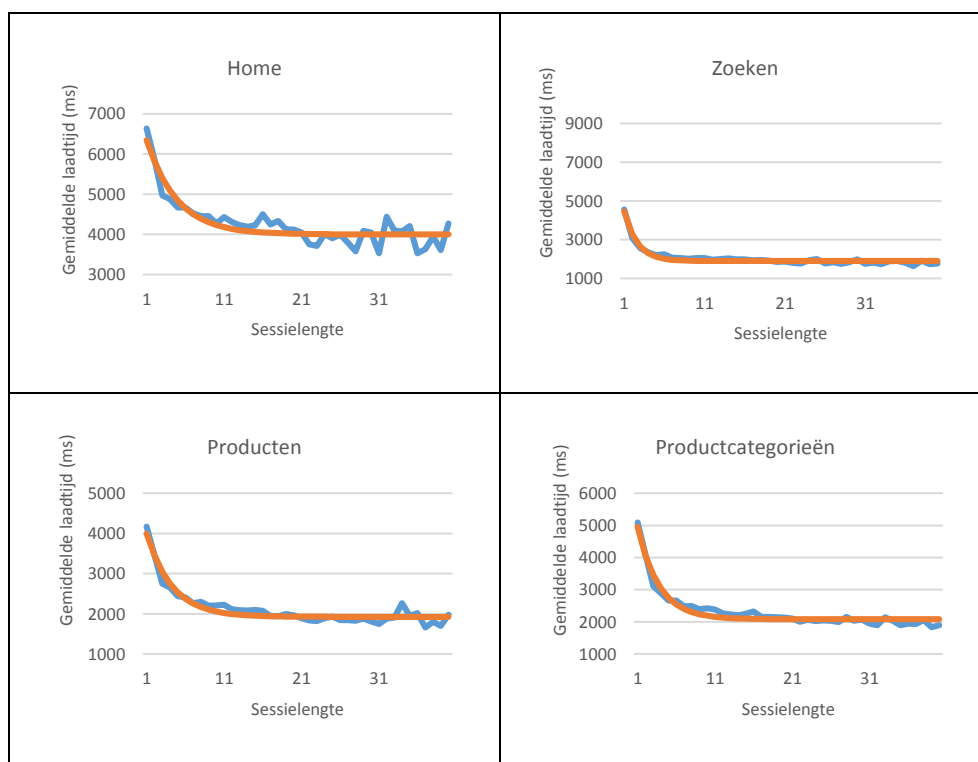
x - de sessielengte

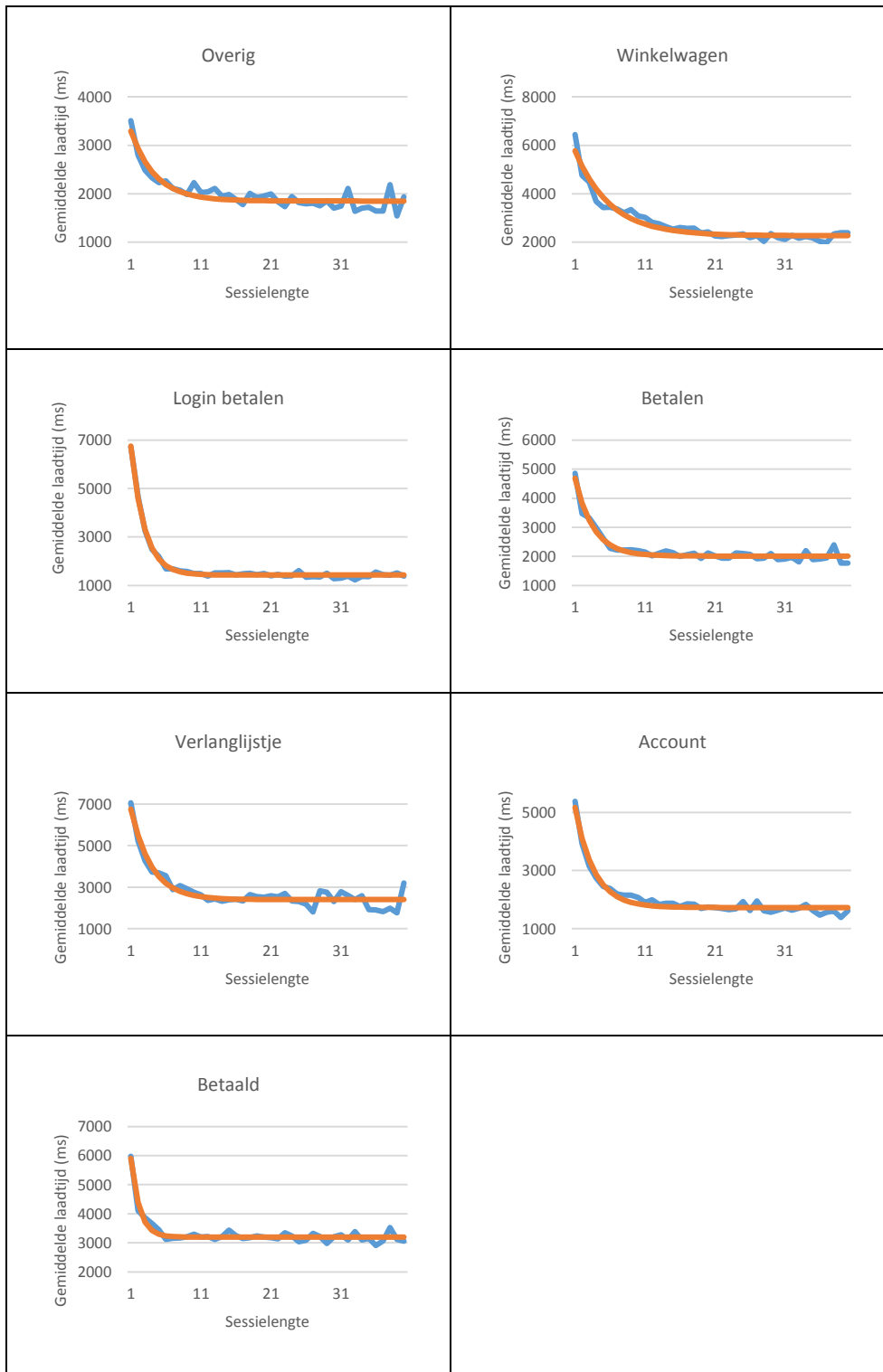
a – *spreidingsparameter*: de spreiding van gemiddelde laadtijd; hoe hoger dit getal hoe groter de spreiding van de laattijden per sessie-lengte

b – *intensiteitsparameter*: de steilheid van de grafiek: hoe steiler de grafiek hoe groter het verschil in laadtijd

c – *ondergrens*: de ondergrens van de gemiddelde laadtijd weer: hoe hoger deze waarde hoe hoger de gemiddelde laadtijd

Vergelijking 5.5 – Exponentiële functie als benadering voor de datapunten.





Figuur 5.6 – Overzicht van curve-fitting per paginagroep aan de hand van MATLAB

Het verband lijkt aan te geven dat een hogere laadtijd ervoor dat bezoekers sneller afhaken. Wanneer de gemiddelde laadtijd namelijk afneemt, wordt het aantal pagina's in de sessie groter.

5.2.3 NONLINEAR LEAST SQUARES FITTING: GAUSS-NEWTON-ALGORITHM

Het fitten van de curve van het verband tussen de gemiddelde laadtijd en sessielengte is gebaseerd op Nonlinear Least Square Fitting. Hierbij wordt er gefit op basis van m observaties met n onbekende parameters waarbij $m \geq n$. In de eerste instantie wordt gebruik gemaakt van een lineaire benadering van de curve. Vervolgens wordt er middels iteratieve stappen getracht de oplossing te verbeteren.

Er zijn meerdere methodes voor het oplossen van 'Nonlinear-Least-Squares-problemen, zoals het Gauss-Newton-algoritme en QR-decompositie. In dit onderzoek is gebruik gemaakt van het Gauss-Newton. Dit algoritme wordt in deze paragraaf beschreven.

5.2.3.1 DE BESCHRIJVING VAN HET GAUSS-NEWTON ALGORITME

Het Gauss-Newton-algoritme gaat iteratief op zoek naar een (lokaal) minimum van de som van de kwadraten van de m residuen door het aanpassen van λ , een vector van n parameters. (MathWorks (2014), Wolfram Alpha (2014)). De parameters in de vector λ zullen door het algoritme worden geoptimaliseerd middels k iteraties. De k -de iteratie vindt plaats op basis van Vergelijking 5.16 in Tabel 5.2. Hoe deze vergelijking tot stand is gekomen wordt beschreven in Vergelijking 5.6 tot en met Vergelijking 5.15 en toegelicht in de volgende alinea. Hierbij zijn de waarden tussen de haakjes de vergelijkingen waarnaar gerefereerd wordt.

Eerst worden de te bepalen parameters gedefinieerd (5.6). Vervolgens dienen deze parameters als input voor de functie y die wordt bepaald aan de hand van de fit (5.7). De observaties (metingen sessielengte vs. Laadtijd) worden benoemd in \hat{y} (5.8). Het residu, oftewel het verschil tussen de observatie en de functie y wordt opgevangen in r (5.9). De som van de kwadratische residuen wordt weergegeven in (5.10). Het doel van het Gauss-Newton algoritme is het minimaliseren van de deze som (5.11). Gedurende elke iteratiestap kan λ worden omschreven middels een eerst-orde-Taylor-reeks (5.11 en 5.12 en uiteindelijk 5.13). In de Taylor-reeks wordt gebruik gemaakt van de Jacobi-matrix (5.14). Dit is een matrix met alle eerste-orde partiële afgeleiden. Wanneer nu de minimalisatie van S uit (5.11) verder wordt uitgevoerd ontstaat (5.15). Het invullen van de uitkomst van deze stap leidt tot de vergelijking in (5.16) die de iteratiestap van λ naar λ^{k+1} beschrijft.

Tabel 5.2 – Beschrijving van het Gauss-Newton algoritme.

Vergelijking	Beschrijving	Formule
5.6	parameters	$\lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$
5.7	functiewaarden	$y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$ waarbij $y_x = f(x, \lambda)$
5.8	observaties	$\hat{y} = \begin{pmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_m \end{pmatrix}$
5.9	residu	$r = y - \hat{y}$
5.10	som van de kwadratische residuen	$S(\lambda) = \ r\ ^2$
5.11	error	$\min_{\lambda} S = \frac{\partial S}{\partial \lambda} = 2r \frac{\partial r}{\partial \lambda} = 0 \quad (\text{voor } j = 1, \dots, n)$

5.12	Taylor-reeks rond λ^k	$f(x, \lambda) \approx f(x, \lambda^k) + J^T \Delta \lambda$
5.13	iteratiestap van λ	$\lambda^{k+1} = \lambda^k + \Delta \lambda$
5.14	Jacobi-matrix	$J = \frac{\partial(r_1, \dots, r_m)}{\partial(\lambda_1, \dots, \lambda_n)} = \begin{bmatrix} \frac{\partial(r_1)}{\partial(\lambda_1)} & \dots & \frac{\partial(r_1)}{\partial(\lambda_n)} \\ \vdots & \ddots & \vdots \\ \frac{\partial(r_m)}{\partial(\lambda_1)} & \dots & \frac{\partial(r_m)}{\partial(\lambda_n)} \end{bmatrix}$
5.15	minimalisatie	$\frac{\partial S}{\partial \lambda} = -2J^T(r - J\Delta\lambda) = 0$ $(J^T J)\Delta\lambda = J^T \Delta y$ $\Delta\lambda = (J^T J)^{-1} J^T \Delta y$
5.16	iteratiestap van λ	$\lambda^{k+1} = \lambda^k + (J^T J)^{-1} J^T \Delta y$

5.2.3.2 TOEVOEGING OP ALGORITME: FACTOR WEGING

Wanneer er een extra factor W als weging wordt gebruikt om een dataset naar een functie te fitten, worden de vergelijkingen om S en λ^{k+1} te berekenen aangepast naar de vergelijking weergegeven in Vergelijking 5.17. Hierbij is W een diagonale matrix met daarin de gewichten van de datapunten. Op deze manier is het mogelijk de residuen van bepaalde datapunten zwaarder te laten meewegen bij het bepalen van de variabelen λ .

$$S = W||R||^2$$

$$\lambda^{k+1} = \lambda^k + (J^T W J)^{-1} J^T W r$$

Vergelijking 5.17 – Het Gauss-Newton-algoritme voor een fit met weging van de datapunten.

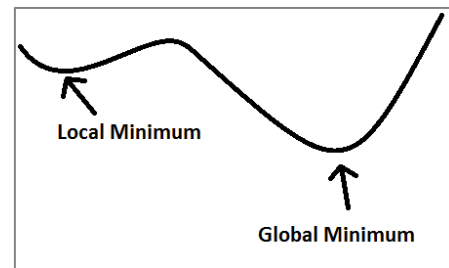
5.2.3.3 CONVERGENTIE VAN HET ALGORITME

De aanpassing van λ gedurende iedere iteratie leidt er altijd toe dat de S afneemt. Wanneer het algoritme convergeert dan zal er altijd een stationair punt door de λ worden bereikt. Wanneer het algoritme niet convergeert, zal er geen oplossing worden gevonden.

Het kan ook voorkomen dat een algoritme zeer langzaam convergeert. In dit geval is er een stopconditie noodzakelijk. Een voorbeeld van zo'n stopconditie kan zijn wanneer geldt dat $\Delta\lambda \leq 0.001$.

5.2.3.4 INITIALISATIE VAN DE COEFFICIËNTEN

Zoals beschreven is er een initiële waarde vereist om te kunnen starten met de NonLinear Least Squares Fitting op basis van het Gauss-Newton-algoritme. Omdat het algoritme kan convergeren in een lokaal minimum is het van belang de initiële waarde zo te kiezen dat het zeer waarschijnlijk is dat het algoritme naar een globaal minimum convergeert.



De initiële waarde wordt via een Random-Brute-Force-Search-methode (RBFS-methode) berekend (Harder, Onbekend). Op die manier is het mogelijk een redelijk goede schatting in de buurt van het globale minimum te geven. Wanneer de initiële waarde is gezet, zal de Gauss-Newton-algoritme de coëfficiënten verder optimaliseren.

Het volgende proces duurt 400 iteraties lang. Hierbij wordt de λ gekozen op basis van het minimaliseren van $S(\lambda)$ en vervangen indien er een oplossing wordt gevonden met een lagere $S(\lambda)$:

De RBFS-methode start met een waarde voor λ middels een schatting van a en c en neemt voor b de waarde 0.75. Vervolgens wordt de $S(\lambda)$ berekend. Vanaf dat moment worden de coëfficiënten met een toevals-waarde (binnen een zekere range) aangepast. Wanneer de aanpassing leidt tot een verbeterde combinatie, dus een lagere $S(\lambda)$, wordt λ aangepast. Wanneer er vervolgens achtereenvolgens twee iteraties zijn waarbij geen verbetering wordt getroffen in $S(\lambda)$, zal er een nieuwe combinatie van coëfficiënten worden gegenereerd aan de hand van een random-factor. Hierbij is de mate van toeval groter dan in de eerder genoemde situatie. Ook in dit geval wordt gekeken of dit een verbetering is ten opzichte van de tot nu toe beste waarde. De iteraties stoppen wanneer een x -aantal iteraties heeft plaatsgevonden.

Aan het einde van dit proces is er een initiële waarde voor λ bepaald die als input dient van het Gauss-Newton-algoritme.

5.2.3.5 ORC₂-TOEPASSING

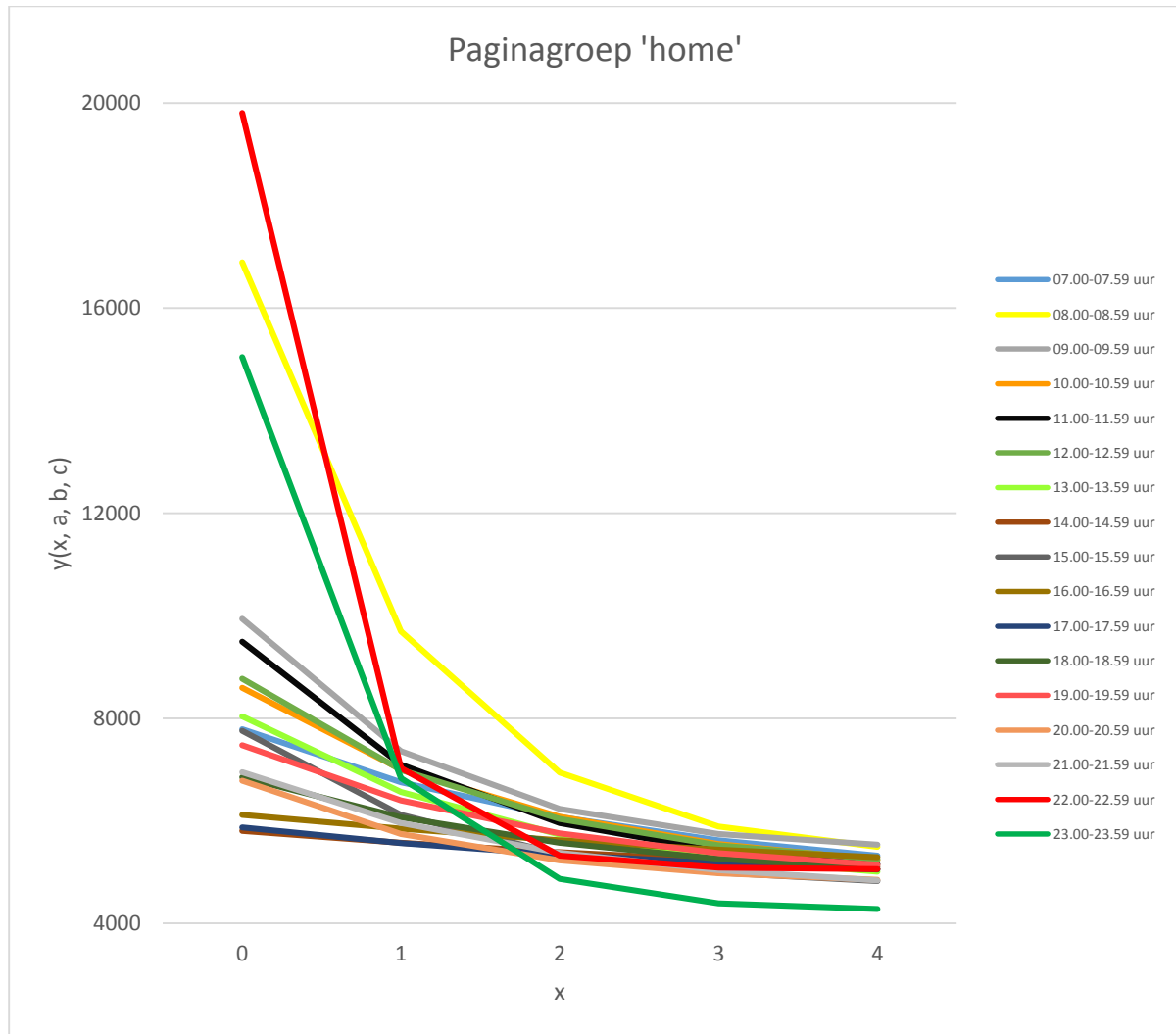
Wanneer het algoritme wordt toegepast om de ORC₂-waarde berekenen geldt dat λ een vector wordt van de n parameters a , b en c zoals beschreven in Vergelijking 5.5 en m is het aantal observaties (verschillende sessielengtes). De gewichten die worden gebruikt, zijn gebaseerd op de frequentie van het aantal pageviews per sessielengte. Op deze manier wordt de observatie gebaseerd op meerdere metingen, zwaarder meegewogen tijdens het vinden van de optimale functieparameters.

Omdat het aantal waarnemingen van een sessielengte groter dan 21 zo klein is dat er weinig rekening mee wordt gehouden gedurende de parametersoptimalisatie, is besloten de optimalisatie alleen voor pageviews waarvan de sessielengte kleiner dan of gelijk is aan 20, uit te voeren.

De ORC₂-waarde wordt per paginagroep berekend. Wanneer er geen onderscheid zou worden gemaakt op basis van het type pagina, zou dit leiden tot een hogere variatie in laadtijd. Als er per paginagroep wordt gekeken is de spreiding van de laadtijd lager omdat de techniek achter de verschillende pagina's in een dergelijke groep in grote mate overeenkomt.

5.2.4 VARIATIE VAN DE FUNCTIE OVER LANGERE PERIODE

In Figuur 5.7 worden de verschillende gefitte-functies weergegeven voor de paginagroep 'home' met als input data gegroepeerd per uur. Opvallend zijn hierbij de functies van de uren 08.00u-08.59u, 22.00u-22.59u en 23.00u-23.59u. Deze lijnen hebben een grotere negatieve richtingscoëfficiënt dan de andere lijnen wat impliceert dat de afname in de laadtijd per extra pageview in een sessie groter is. De conversie-cijfers laten zien dat de gemiddelde conversie op deze momenten negatieve pieken laten zien.



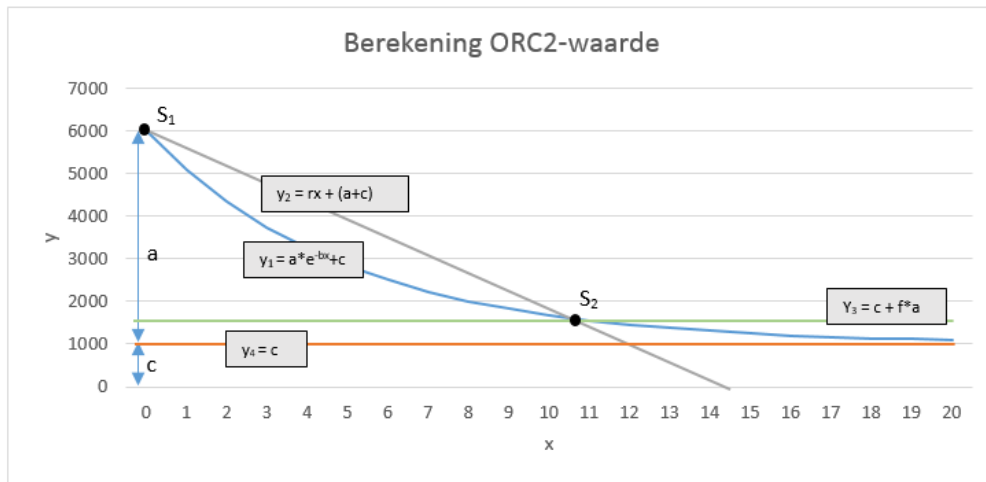
Figuur 5.7 – Datapunten van de paginagroep 'home' gefit aan de hand van de exponentiële functie uit Vergelijking 5.5.

5.2.5 DEFINITIE VAN DE ORC_2

De waarde van de parameters in de vector λ vertellen iets over het verband tussen de gemiddelde laadtijd en de sessielengte. Het signaal van performance-degradatie zal door de ORC_2 -waarde moeten worden gegenereerd. Uit Figuur 5.7 valt, samen met de conversie-kennis, af te leiden dat er tussen 08.00u-08.59u en 22.00u-23.59u een negatieve verandering van performance plaatsvindt dat leidt tot een afname van de conversie. Hoe kan deze verandering worden opgemerkt? Met andere woorden, wat is de definitie van de ORC_2 ?

De ORC_2 wordt bepaald door de steilheid van de grafiek te berekenen, gebruikmakend van alle parameters. Elk van deze parameters bevat namelijk belangrijke informatie over de spreiding, intensiteit en ondergrens van de functie. De ORC_2 is daarom gebaseerd op de richtingscoëfficiënt van de rechte lijn door de verticale as (S_1) en een punt net boven de horizontale asymptoot (S_2), zoals weergegeven in Figuur 5.8. Dit laatste punt wordt

gedefinieerd door: $c + f \cdot a$ en wordt in de volgende alinea nader toegelicht. De coëfficiënt zegt iets over de steilheid van de grafiek en daarmee iets over verhouding van de sessielengte en de gemiddelde laadtijd.



Figuur 5.8 - Berekening van de ORC₂

Omdat de ondergrens nooit bereikt zal worden (asymptoot), is het noodzakelijk te kijken naar een waarde net boven de ondergrens om zo een maat te berekenen voor de absolute steilheid van de grafiek. Deze waarde ligt op de $y = c + f \cdot a$, waarbij f een factor $\in (0,1)$.

Vervolgens wordt de richtingscoëfficiënt m bepaald van de lijn $y_2 = rx + (a+c)$. Deze lijn heeft punten op S_1 en op S_2 . Dit zijn de snijpunten van de lijnen y_1 met y_2 en de lijnen y_1 met y_3 .

De richtingscoëfficiënt r wordt berekend door Vergelijking 5.18.

$$r = \frac{\Delta y}{\Delta x} = \frac{(c + fa) - (a + c)}{\frac{\ln(f)}{-b}} = \frac{a(f - 1)}{\frac{\ln(f)}{-b}} = -\frac{ab(f - 1)}{\ln(f)} = \frac{ab(1 - f)}{\ln(f)}$$

Vergelijking 5.18 – Berekening van de richtingscoëfficiënt van de lijn door S_1 en S_2 .

Waarbij Δx gevonden wordt door het oplossen van $y_1 = y_3$ (Vergelijking 5.19).

$$\begin{aligned} ae^{-bx} + c &= c + fa \\ ae^{-bx} &= fa \\ -bx &= \ln(f) \\ x &= \frac{\ln(f)}{-b} \end{aligned}$$

Vergelijking 5.19 – Het bepalen van Δx .

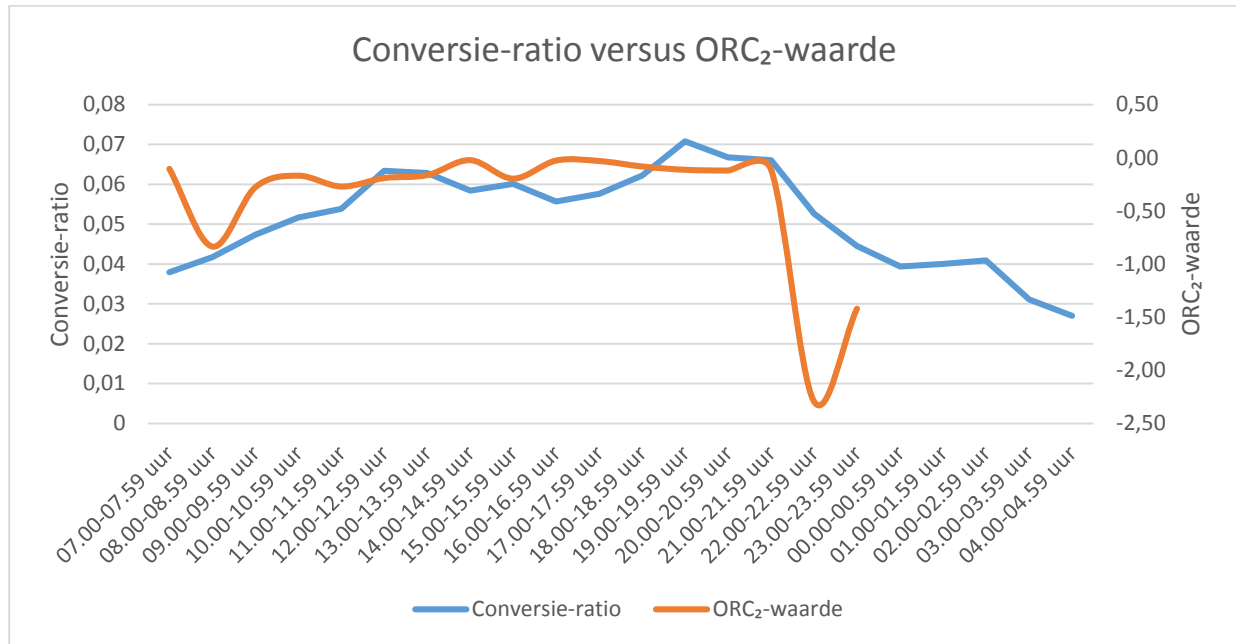
De ORC₂ is op dit moment slechts afhankelijk van a en b en wordt daarom met parameter c geschaald. De definitie van de ORC₂ is als weergegeven in Vergelijking 5.20. Hierbij dient de f als logaritmische schalingsfactor.

$$ORC_2 = \frac{r}{c} = \frac{\frac{ab(1 - f)}{\ln(f)}}{c} = \frac{ab(1 - f)}{c \cdot \ln(f)}$$

Vergelijking 5.20 - Definitie ORC₂-waarde.

5.2.6 SIGNALERING VAN PERFORMANCE-DEGRADATIE DOOR ORC2

Figuur 5.9 geeft de conversie-ratio en ORC₂-waardes per van data per uur van de paginagroep 'home'. De exponentiële functies zijn weergegeven in Figuur 5.7. Wat opvalt, is dat wanneer de ORC₂-waarde zeer afneemt (van 21.00u-21.59u tot 22.00u-22.59u), de conversie-ratio ook lijkt af te nemen. Dit zou veroorzaakt kunnen worden door er mogelijk een dag- en nacht-afhankelijkheid bestaat bij de conversie-ratio. Echter lijkt de daling van de conversie-ratio zich wel dusdanig sterk in te zetten dat het lijkt alsof er een verband tussen de twee variabelen bevestigd wordt.



Figuur 5.9 – Conversie-ratio versus de ORC₂-waarde van de dataset in Figuur 5.6.

De waarde voor de ORC₂ wordt interessant wanneer het verloop van deze waarde per paginagroep door de tijd heen wordt beschouwd, zoals in de figuur getoond. De werkelijke ORC₂-waarde lijkt in eerste instantie niet van groot belang: de variatie van de ORC₂ per paginagroep verdient de boventoon.

6 IMPLEMENTATIE METHODIEK

De in Hoofdstuk 5 beschreven methodiek is geïmplementeerd in PHP om zo direct op de tag aangesloten te kunnen worden. Ook zijn er tabellen gecreëerd waarin de resultaten van de methodiek kunnen worden opgeslagen.

6.1 OPSLAG METHODIEK

Om de methodiek te kunnen uitvoeren is het noodzakelijk bepaalde data op te kunnen slaan. Naast de tabel die is gemaakt om de algemene performance-data van de tag op te slaan, zijn er een aantal andere tabellen ontworpen (Tabel 6.1).

Tabel 6.6.1 – Overzicht tabellen om de methodiek te kunnen uitvoeren (m staat voor methodiek).

Tabelnaam	Toelichting
m_notification	Overzicht van alle sessies waarnaar een notificatie is gestuurd.
m_orc1	Overzicht van alle ORC ₁ -waardes van alle paginagroep-combinaties.
m_orc2	Overzicht van de huidige ORC ₂ -waardes per paginagroep.
m_orc2_history	Overzicht van alle ORC ₂ -waardes van de afgelopen vanaf middernacht.
m_overzicht	Geaggregeerde dag-data per domein per paginagroep.
m_pagelabels	Overzicht van de gebruikte paginagroepen per domein.
m_pageviews	Overzicht van alle pageviews van de afgelopen drie uur inclusief methodiek-waardes.
m_sessionlength	Overzicht van het aantal sessies per sessielengte in combinatie met gemiddelde laadtijd.

6.2 ORC₁

De ORC₁-waarde is gebaseerd op wijze van navigatie tussen paginagroep door de converterende klanten. Omdat de wijze van navigatie redelijk stabiel lijkt te blijven gedurende een langere periode, wordt de frequentie van het berekenen van de ORC₁-waarde laag ingesteld. Er zal dagelijks om middernacht een update plaatsvinden.

6.3 ORC₂

Voor het bepalen van de ORC₂-waarde is er veel data vereist. Wanneer er een beperkt aantal pageviews is geregistreerd is het niet mogelijk binnen een zekere foutmarge een exponentiële functie op te stellen voor de ORC₂-waarde. In dit geval is het een optie om de periode waarover de ORC₂-waarde wordt bepaald te verruimen: waar in eerste instantie een ORC₂-waarde per vijf minuten zou worden opgesteld, is er nu voor gekozen de ORC₂-waarde op uur-basis op te stellen.

De berekening van de coëfficiënten van de exponentiële verdeling gebeurt onder andere via Gauss-Newton. Als basis is hiervoor gebruik gemaakt van een al bestaande PHP-class (Que, 2013). Deze is aangepast om de gewichten er in te kunnen verwerken

6.4 WEERGAVE VAN DE RESULTATEN

Om de methodiek te kunnen verifiëren en valideren is er een visuele weergave van de methodiek gemaakt. Deze is momenteel te vinden op de beacon.mwtag.nl/MeasureWorksTag.php. Op deze manier is inzichtelijk hoe de ORC₁ en ORC₂ zich gedurende een periode gedragen.

7 RESULTATEN

Om de in Hoofdstuk 5 beschreven methodiek te kunnen testen, is het noodzakelijk performance- en conversie-data te vergaren. Dit bleek uiteindelijk moeilijker dan in eerste instantie gedacht: klanten van MeasureWorks die hadden toegezegd, bleken toch terug te krabbelen of een implementatietijd te kennen die de resterende tijd van de stage overtrof. Dit is ook de reden dat de test uit Hoofdstuk 8 is opgezet.

Toch is het gelukt de tag in een zestal websites te plaatsen. Helaas gaat het hierbij niet om werkelijke e-commerce sites met hoge bezoekersaantallen, waardoor de methodiek niet volledig tot zijn recht kan komen. De methodiek wordt daarom deels aangepast door een grotere periode te hanteren bij het berekenen van de ORC₁ en de ORC₂: deze worden nu gebaseerd op een week aan data.

7.1 WOLFRIJBEWIJSSHOP.NL

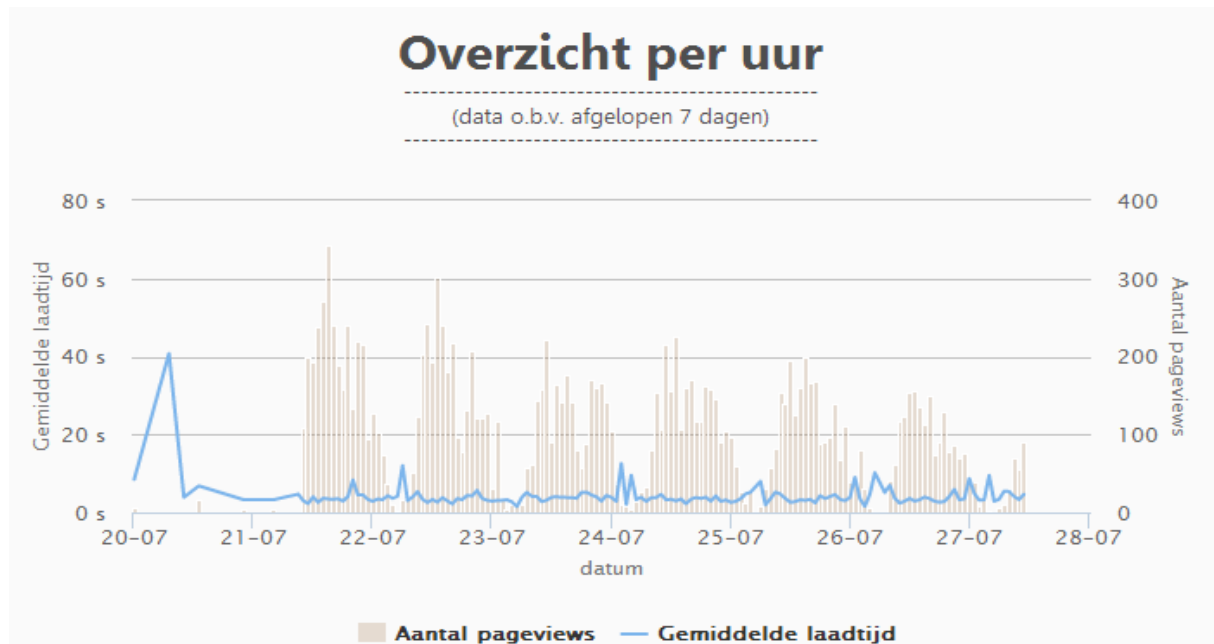
De website, waarbij de tag is geïmplementeerd, met de hoogste bezoekersaantallen is www.wolfrijbewijsshop.nl (Figuur 7.1). Deze website heeft ongeveer 2500-3000 pageviews per dag en heeft als conversie-doel het maximaliseren van het aantal door-bezoekers-ingevulde-web-formulieren. Dit betekent dat dit geen e-commerce website betreft.

The screenshot shows the homepage of wolfrijbewijsshop.nl. At the top left is the logo 'WOLF rijbewijsshop.nl'. To its right is a counter 'Tot nu toe 024399 tevreden geslaagden' and a phone number 'Bel 088 1000 800 (lokaal tarief)'. A navigation bar contains links: 'Rijbewijs halen', 'Rijlessen Plannen', 'Video's', 'Overstapservice', 'Over WOLF Rijbewijsshop', 'Contact', and 'Direct inschrijven!'. The main heading is 'Wil je ook snel slagen voor je rijbewijs?' followed by 'Meer dan 20.000 zijn je voor gegaan, omdat...'. A list of benefits includes: 'Altijd in de buurt', 'Minder lessen door unieke lesmethode', 'Gespreid betalen', 'Hoog slagingspercentage', and 'Juiste advies & direct starten'. A yellow button says 'Ik wil ook slagen!'. Below this is a carousel with three items: 'Spreek een Intake af!', 'Snel je theoriecertificaat!' (with a price of €69,- and 'OP=OP!' label), and 'Haal direct je rijbewijs!'. The background features a woman celebrating next to a car with 'WOLF' on it.

Figuur 7.1 – Homepagina van wolfrijbewijsshop.nl.

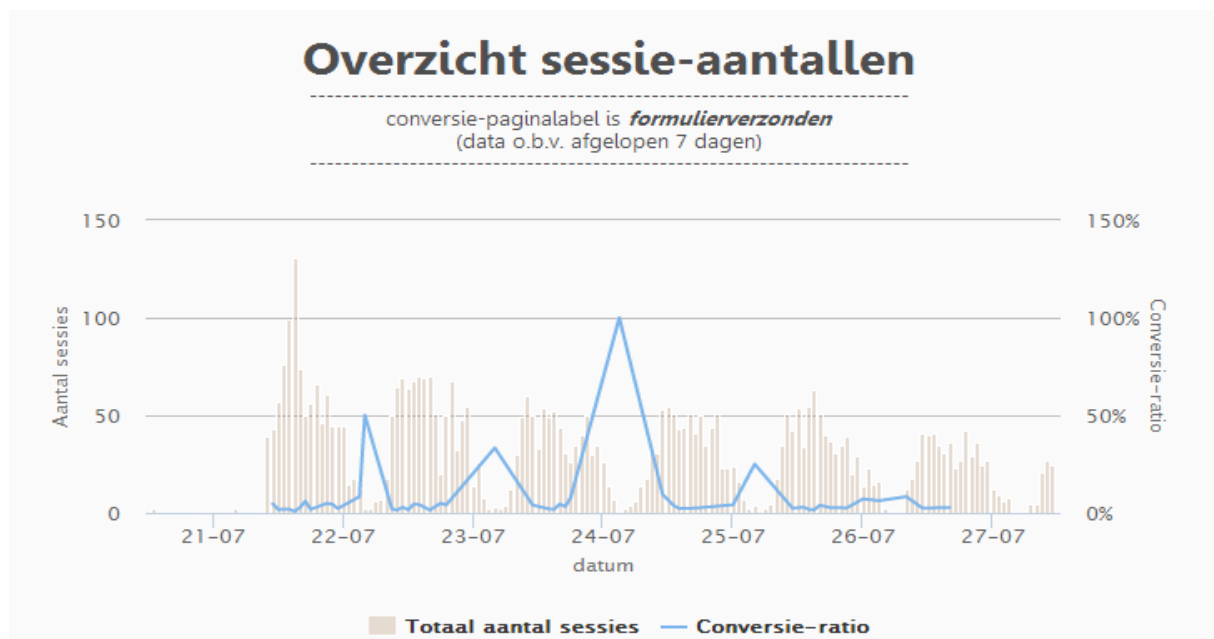
Naast het feit dat deze website in vergelijking met grote e-commerce sites lage bezoekersaantallen heeft, is ook het conversie-doeleinde verschillend. Bezoekers kunnen op veel meer pagina's overgaan tot conversie: meer dan de helft van de pagina's biedt de mogelijkheid om een contactformulier in te vullen.

De resultaten die volgen hebben betrekking op webpagina die is ontwikkeld ter visualisatie van de methodiek (Kooiman, 2014). In Figuur 7.2 is een overzicht te vinden van een momentopname van het aantal pageviews op wolfrijbewijsshop.nl. Een technisch probleem met de MW-tag verklaart hierbij het deels ontbreken van data op 20 en 21 juli. Ook is te zien dat naarmate een week vordert, het aantal pageviews afneemt.



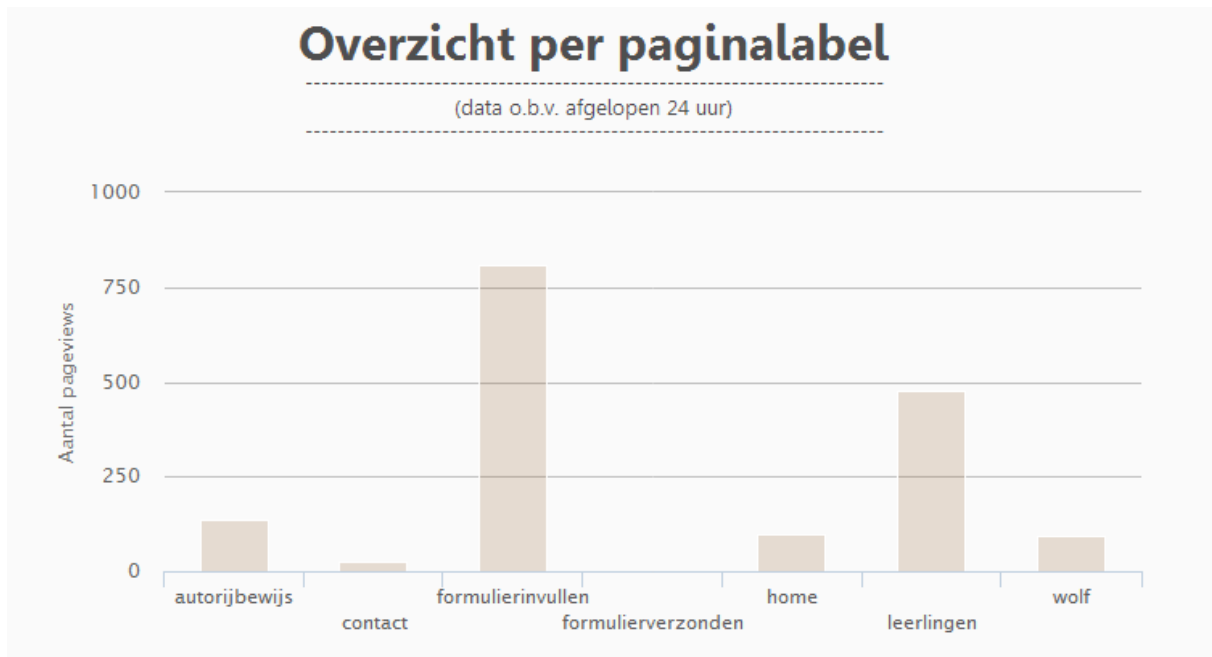
Figuur 7.2 – Aantal pageviews en gemiddelde laadtijd van wolfrijbewijsshop.nl in de periode van 20 juli 2014 – 28 juli 2014.

Figuur 7.3 geeft een overzicht van de bezoekersaantallen en de conversie-ratio. De conversie-ratio is berekend door eerst te conversie-paginagroep te definiëren en vervolgens het aantal sessies dat deze paginagroep bezoekt als ratio van het totaal aantal sessies gebruiken. In het geval van wolfrijbewijsshop.nl is gekozen om de paginagroep *formulierverzonden* als conversie-paginagroep te benoemen. Zoals eerder genoemd is namelijk het aantal ingevulde web-formulieren een maat voor de conversie van deze website. De conversie-ratio vertoont rond middernacht pieken, omdat er dan weinig bezoekers zijn waardoor de conversie-ratio veel volatieler is.



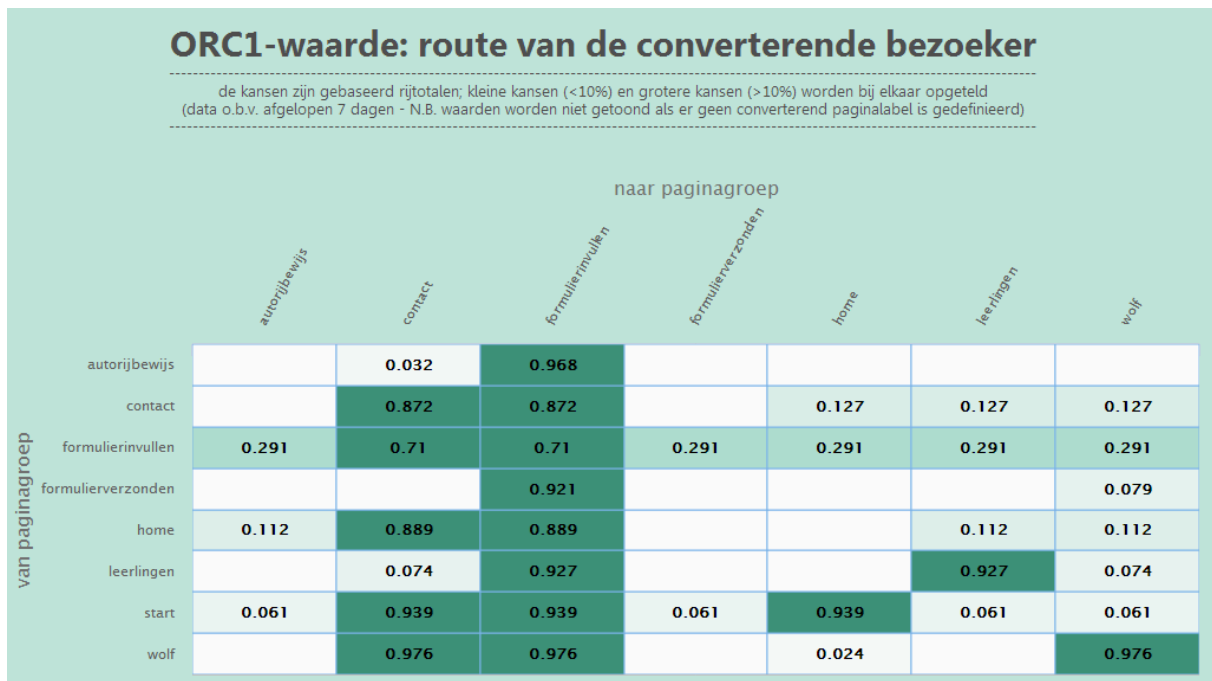
Figuur 7.3 - Overzicht van de sessie-aantallen en conversie-ratio van wolfrijbewijsshop.nl in de periode van 20 juli 2014 – 28 juli 2014.

Figuur 7.4 geeft een overzicht van pageviews per paginagroep. Hierbij valt op dat de meeste pagina's die worden bezocht in de paginagroep *formulierinvullen* of *leerlingen* vallen. Dit eerste is te verklaren doordat er zoals gezegd zeer veel pagina's zijn waarop een formulier kan worden ingevuld. Naast de mogelijkheid om een formulier in te vullen wordt er ook andere informatie aangeboden met betrekking tot de rijkschool.



Figuur 7.4 – Overzicht aantal pageviews per pagelabel periode 26 juli 2014 13.00u – 27 juli 2014 13.00u wolfrijbewijsshop.nl.

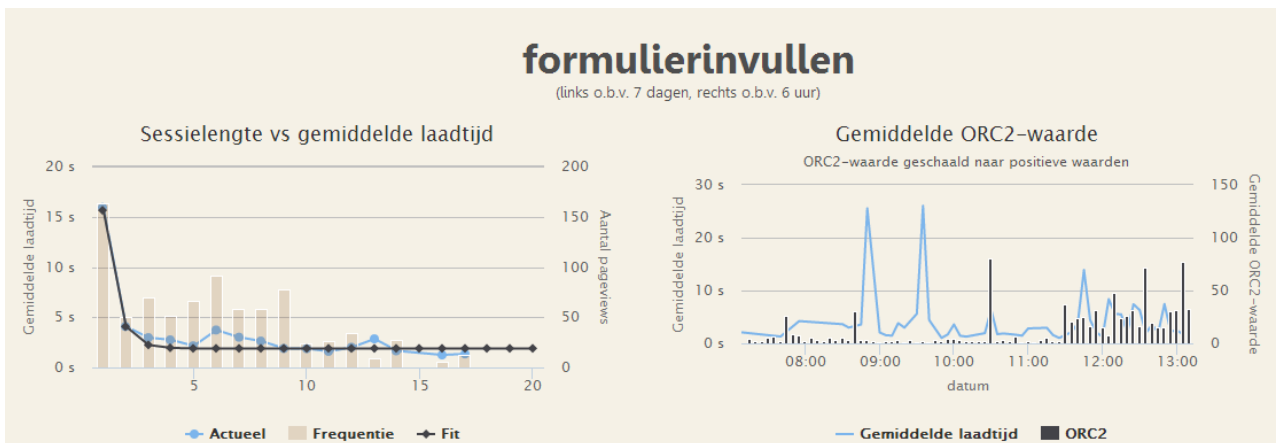
Figuur 7.5 geeft een overzicht van de ORC₁-waardes. Hierbij is duidelijk te zien dat er per tabelrij twee mogelijke waardes zijn te onderscheiden. Opvallend is dat bezoekers naar de paginagroep *formulierinvullen* navigeren ongeacht de voorgaande pagina. Dit wordt veroorzaakt doordat relatief veel bezoekers een pagina van de paginagroep *formulierinvullen* bezoeken.



Figuur 7.5 – ORC₁-waarde wolfrijbewijsshop.nl op basis van 20 juli 2014 13.00u - 27 juli 2014 13.00u.

Figuur 7.6 geeft de grafieken met informatie over de ORC_2 -waarde weer. De linker-grafiek geeft de beste fit weer op basis van de RBFS-methode en het Gauss-Newton-algoritme (Hoofdstuk 5) en de rechter-grafiek geeft de ORC_2 -waarde en gemiddelde laadtijd door de loop van de tijd heen, weer. Belangrijk te vermelden is dat de ORC_2 -waarde in spiegelbeeld ten opzichte van de horizontale as is weergegeven. Dit maakt de grafiek beter leesbaar dan wanneer de laadtijd positief is en de ORC_2 -waarde negatief.

Opvallend is dat de ORC_2 -waarde toeneemt tussen 11.00u en 12.00u. Op die momenten is ook te zien dat de gemiddelde laadtijd toeneemt. Belangrijk voor het interpreteren van de data is dat de linker-grafiek iedere vijf minuten op basis van 24 uur aan historische data wordt opgesteld. Dit betekent dat er een zeer grote overlapping van datapunten is bij iedere nieuwe ORC_2 -berekening (meer dan 99%). Echter, met minder datapunten is de ORC_2 -waarde niet betrouwbaar omdat de exponentiële vergelijking dan een te grote foutmarge ten opzichte van de dataset kent.



Figuur 7.6 – ORC_2 -waarde wolfrijbewijsshop.nl op basis van 20 juli 2014 13.00u - 27 juli 2014 13.00u.

De ORC_2 -waarde waarbij kan worden vastgesteld dat er een signaal moet worden afgegeven naar de bezoeker zal per site afhankelijk zijn. Het gaat met name om de fluctuatie in de ORC_2 -waarde in verloop van tijd: de uitschieters moeten worden gesignaleerd.

8 TEST: HOE REAGEREN BEZOEKERS OP BERICHTGEVING BIJ VERTRAGINGEN IN EEN WEBSHOP?

Omdat het in praktijk lastig is gebleken om binnen de gestelde termijn een grote webshop bereid te vinden de tag te implementeren, is er op zoek gegaan naar een manier om te testen wat de informatievoorziening in praktijk zou opleveren. Er is daarom een test ontwikkeld waarin deelnemers is gevraagd een artikel in een versimpelde versie van een webshop te kopen. De deelnemers zijn ingedeeld in een van de zeven opgestelde scenario's en kregen afhankelijk van dit scenario te maken met verschillende laadtijden. Ook is in een aantal scenario's een bericht getoond met betrekking tot de huidige performance om zo mogelijk de performance-tolerantie te verhogen. In dit hoofdstuk worden de exacte opzet van de test, de uitvoering en de verkregen resultaten van de test beschreven. Ook wordt de conclusie van dit sub-onderzoek gepresenteerd.

8.1 DOEL VAN DE TEST

Het doel van de test is het onderzoeken of berichtgeving bij trage pagina's er voor zorgt dat er minder bezoekers afhaken. Er wordt hierbij gebruik gemaakt van zeven verschillende scenario's. Aan de hand van deze scenario's zijn acht hypothesen opgesteld welke middels Fisher's Exact Test zullen worden getest.

8.2 BESCHRIJVING VAN DE TEST

8.2.1 VOORAFGAAND AAN HET BEZOEK AAN DE WEBSHOP

In de test wordt iedere deelnemer op de eerste pagina gevraagd een Nederlands boek aan te schaffen in een versimpelde webshop. Wanneer deelnemers gedurende hun bezoek aan de webshop om de een of andere reden niet tevreden zijn over de webshop, wordt hen gevraagd op de knop 'verlaat de webshop!' te drukken. Wanneer de deelnemer doorklikt, wordt de homepagina van de webshop geladen.

8.2.2 TIJDENS HET BEZOEK AAN DE WEBSHOP

Omdat het om een versimpelde versie van een webshop gaat, hebben de deelnemers een beperkt aantal navigatie-mogelijkheden binnen de website. Zo is er qua producten alleen een overzicht van de top-tien Nederlandse boeken. Wanneer de productdetails van dit boek worden bekeken is het mogelijk het boek toe te voegen aan het winkelmandje. Wanneer dit gebeurt, kan de deelnemer starten met de tweede fase van het aanschaf-proces, namelijk het daadwerkelijk bestellen van het product.

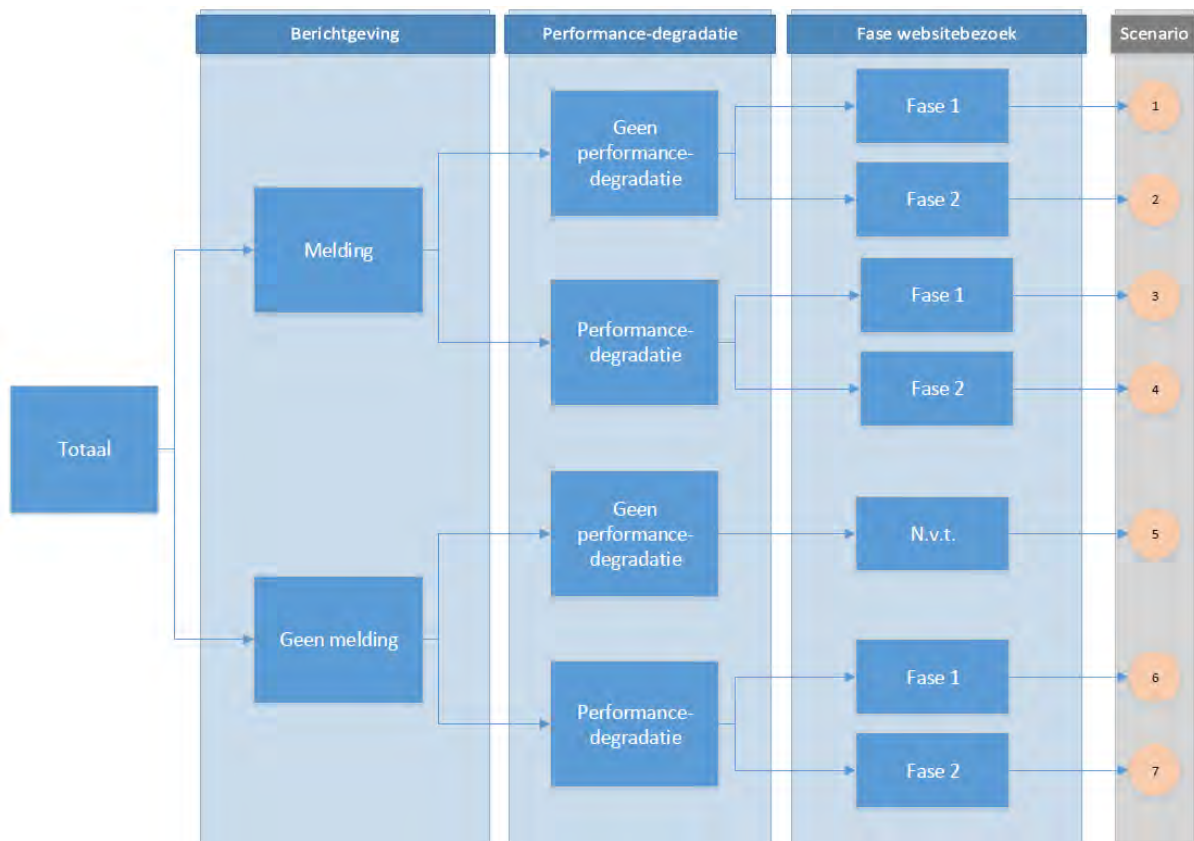
Wanneer de bestelling van een boek is voltooid, worden deelnemers automatisch naar een pagina gestuurd waar vragen met betrekking tot het klikgedrag dienen te worden beantwoord. Ook wanneer bezoekers voortijdig op de knop 'verlaat de webshop!' drukken worden zij automatisch naar een vragenlijst doorgestuurd.

8.2.3 NA HET BEZOEK AAN DE WEBSHOP

Na afloop van het bezoek aan de webshop dient de deelnemer een formulier in te vullen met vragen over de redenen waarom de deelnemer is afgehaakt dan wel tot conversie is overgegaan. In het eerste geval is de knop 'verlaat de webshop' ingedrukt, in het tweede geval is het complete bestelproces voor de aanschaf van een boek doorlopen. Voor beide situaties wordt er een pagina met daarop verschillende vragenlijsten geladen. Beide pagina's bevatten vragen over de keuzes die de deelnemers gedurende hun websitebezoek hebben gemaakt.

8.3 SCENARIO'S

Er is een zevental scenario's opgesteld voor de test om zo de invloed van berichtgeving te kunnen testen. In eerste instantie wordt er onderscheid gemaakt op het al dan niet genereren van een melding. Daarnaast wordt er onderscheid gemaakt op basis van performance: een deel van de bezoekers zal te maken krijgen met hogere laadtijden dan een ander deel. Tot slot is de laatste groepering op basis van de fase waarin men te maken krijgt met de eventuele pagina-vertraging en berichtgeving. Hiervoor is gekozen omdat onderzoek uitwijst dat de fase van het websitebezoek van invloed is op de laadtijd-tolerantie. Figuur 8.1 geeft een overzicht van de verschillende scenario's.



Figuur 8.1 - Overzicht testscenario's

8.3.1.1 FASE VAN HET WEBSITEBEZOEK

Zoals uit Hoofdstuk 2 blijkt, is de fase van het websitebezoek medebepalend voor de performance-tolerantie. Wanneer men zich verder in het conversie-proces bevindt, heeft de performance minder invloed in vergelijking met de situatie waarin een bezoeker zich bijvoorbeeld oriënteert op een product. Daarom is er voor gekozen twee fasen van het websitebezoek te onderscheiden: de oriëntatiefase (fase één) en de bestelfase (fase twee).

8.3.1.2 BERICHTGEVING

Indien er een melding is getoond aan de deelnemer, krijgt de bezoeker het bericht in Figuur 8.2 te zien. Deze melding wordt op de pagina direct voor de eerste pagina met performance-degradatie vertoond. Dit is bij fase één op de pagina met het overzicht van de boeken en bij fase 2 het winkelmandje.

*Beste bezoeker,
 We hebben gemerkt dat de website momenteel minder snel is dan u van ons gewend bent.
 Gelukkig kunt u uw bestelling gewoon blijven plaatsen. Ondertussen wordt er hard aan gewerkt om de
 snelheid weer op peil te krijgen.
 Bedankt voor uw begrip!*

Figuur 8.2 – Berichtgeving ten tijde van performance-degradatie.

8.3.1.3 PERFORMANCE-DEGRADATIE

De performance-degradatie vindt plaats op twee opeenvolgende pagina's. De vertraging die is ingebouwd op de eerste van deze pagina's betreft zes seconden. De vertraging op de tweede pagina betreft tien seconden. De vertraging wordt opgebouwd om de kans te vergroten dat de bezoeker de vertraging al dan niet bewust opmerkt.

Indien er vertraging in het scenario wordt ingezet, zal de eerste vertraging plaatsvinden op de pagina waar ook de mogelijke melding wordt gedaan. Dit is dus voor fase één de pagina met het overzicht van de boeken en voor fase twee is dit de pagina met het winkelmandje.

8.4 HYPOTHESEN

Op basis van de zeven scenario's zijn er acht hypothesen opgesteld. Hierbij wordt een deelnemer als afhaker bestempeld wanneer hij op de knop 'verlaat de webshop!' heeft. Wanneer het gehele bestelproces wordt doorlopen en het boek wordt aangeschaft, wordt het proces als afgerond gezien.

Een belangrijke ratio om te bepalen in welke mate deelnemers het bestelproces niet voltooien tijdens de test, is de afhakers-ratio R_i . Deze wordt gedefinieerd in Vergelijking 8.1. Op basis van deze ratio worden de hypothesen opgesteld zoals weergegeven in Tabel 8.1.

$$R_i = \frac{\text{\#aantal afhakers scenario } i}{\text{\#totaal aantal observaties scenario } i} \quad i = (1, \dots, 7)$$

Vergelijking 8.1 – Definitie R_i .

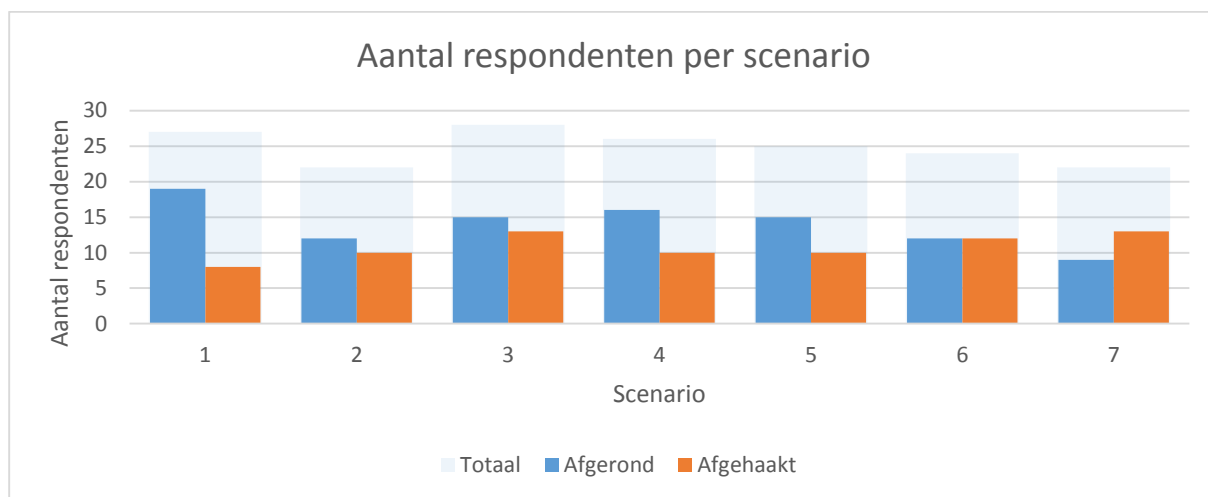
Tabel 8.8.1 – Overzicht hypothesen.

Individuele scenario's	H_0	H_1
1a. Scenario 1 kent significant minder afhakers dan scenario 3.	$R_1 \geq R_3$	$R_1 < R_3$
1b. Scenario 2 kent significant minder afhakers dan scenario 4.	$R_2 \geq R_4$	$R_2 < R_4$
2a. Scenario 3 kent significant minder afhakers dan scenario 6.	$R_3 \geq R_6$	$R_3 < R_6$
2b. Scenario 4 kent significant minder afhakers dan scenario 7.	$R_4 \geq R_7$	$R_4 < R_7$
Gegroepeerde scenario's	H_0	H_1
3. Scenario 5 kent significant minder afhakers dan scenario 6 en 7.	$R_5 \geq R_{6,7}$	$R_5 < R_{6,7}$
4. Scenario's 1, 2, 3 en 4 kennen significant minder afhakers dan de scenario's 5,6 en 7.	$R_{1,2,3,4} \geq R_{5,6,7}$	$R_{1,2,3,4} < R_{5,6,7}$
5. Scenario's 3 en 4 kennen significant minder afhakers dan de scenario's 6 en 7.	$R_{13,4} \geq R_{6,7}$	$R_{3,4} < R_{6,7}$
6. Scenario's 1, 2 en 5 kennen significant minder afhakers dan de scenario's 3,4,6 en 7.	$R_{1,2,5} \geq R_{3,4,6,7}$	$R_{1,2,5} < R_{3,4,6,7}$

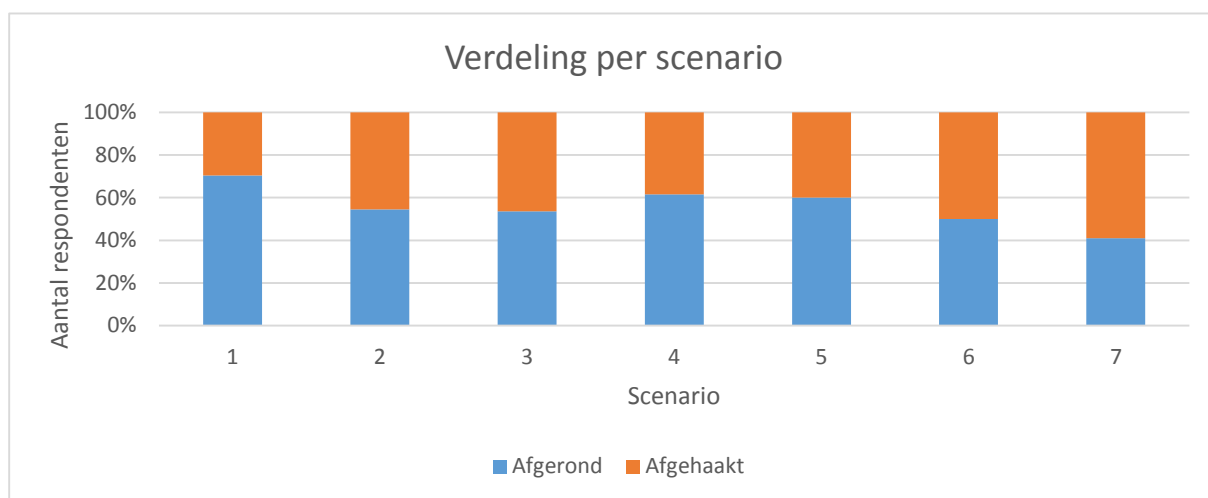
8.5 RESULTATEN

8.5.1 VERDELING RESPONDENTEN: AFGEHAAKT OF AFGEMAAKT

Figuur 8.3 geeft een overzicht van het aantal respondenten per scenario. In totaal hebben 174 respondenten de pagina met daarop een formulier bereikt. Ook wordt de verhouding tussen de tests waarin men is afgehaakt en de tests waarin men de aanschaf van boek heeft voltooid, getoond (Figuur 8.4). Te zien is dat scenario drie het meest vaak is getest door de respondenten. Verder is op te merken dat in scenario één tot en met scenario vijf er meer respondenten de test hebben voltooid dan eerder zijn afgehaakt.



Figuur 8.3 – Aantal respondenten per scenario.



Figuur 8.4 – Verdeling aantal respondenten per scenario.

8.5.2 FISHER'S EXACT TEST

De hypothesen worden getest aan de hand van Fisher's Exact Test (Weisstein, Onbekend). Aan de hand van deze test zal blijken of de uitkomsten van de onderzoek statistisch significant zijn. Deze test is gebaseerd op de hypergeometrische kansverdeling en wordt gebruikt voor het analyseren van kruistabellen. De verdeling is weergegeven in Vergelijking 8.2.

Er worden n trekkingen uit populatie N gedaan zonder teruglegging. In deze populatie zijn er K successen en in de getrokken steekproef zijn er k successen. Hierbij is het aantal successen in deze setting gelijk aan het aantal afhakers in het scenario waarvan de R_i lager wordt geschat onder H_0 .

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

met

- X – aantal succes als stochastische variabele
- K – aantal successen in de populatie
- k – aantal successen in de steekproef
- N – populatiegrootte
- n – steekproefgrootte

Vergelijking 8.2 – Hypergeometrische verdeling

De eenzijdige p-waarde wordt berekend door te sommeren over alle kansen op mogelijkheden waardoor H_0 wordt versterkt. Dit betekent dat er wordt gesommeerd over alle $P(X = i)$ voor alle waarden van $i = 0$ tot en met k (Vergelijking 8.3). De p-waarde onder een gegeven hypothese zegt iets over de waarschijnlijkheid dat een situatie zich voor doet. Als grens wordt hierbij meestal 5% gebruikt: wanneer de p-waarde kleiner is dan 5% betekent dit dat de kans dat een gebeurtenis plaatsvindt zodanig klein is dat de hypothese waarmee de p-waarde is berekend, verworpen kan worden.

Het betreft hier een eenzijdige kans omdat de hypothese niet is gebaseerd op de waarschijnlijk van gelijkwaardigheid van twee waarden, maar op de waarschijnlijkheid dat de ene waarde groter of gelijk aan de andere waarde is.

$$p - \text{waarde} = \sum_{i=1}^k P(X = i)$$

Vergelijking 8.3 - Eenzijdige p-waarde

Wanneer Vergelijking 8.2 en Vergelijking 8.3 op Hypothese 1a van de test wordt toegepast, ontstaan de resultaten weergegeven in Figuur 8.5. Het testen van de Hypothese 1b tot en met Hypothese 6 hebben op een dergelijke wijze plaatsgevonden.

Hypothese	Scenario 1 kent significant minder afhakers dan scenario 3.			
	$H_0: R_1 \geq R_3$ $H_1: R_1 < R_3$			
Aantal waarnemingen		Scenario 1	Scenario 3	Totaal
	Aantal niet-voltooid tests	8	13	21
	Aantal voltooid tests	19	15	34
	Totaal	27	28	55

<i>Parameters</i>	$K = 27$ $k = 8$ $N = 55$ $n = 21$
<i>Hypergeometrische kansberekening</i>	$P(X = 8) = \frac{\binom{27}{8} \binom{28}{13}}{\binom{55}{21}} = 0,0988$
<i>Enkelzijdige p-waarde</i>	<p>Hypergeometrische kansen: $N=55, n=21, K=27$</p> <p>p-waarde = 0,2695</p>
<i>Conclusie</i>	<p>H_0 kan niet verworpen worden. Er is dus geen significant bewijs dat performance-degradatie leidt tot meer afhakers.</p>

Figuur 8.5 – Voorbeeldberekening Fisher's Exact Test.

Tabel 7.2 geeft een overzicht van de geteste hypothesen. Wanneer er een p-waarde van 5% wordt gebruikt om de hypothesen al dan niet te verwerpen, betekent dit dat alle nulhypothesen niet kunnen worden verworpen. Alle p-waardes zijn namelijk groter dan 5%.

Tabel 8.2 – Overzicht hypothesen

Hypothese	H_0	H_1	p-waarde
1a	$R_1 \geq R_3$	$R_1 < R_3$	0,1576
1b	$R_2 \geq R_4$	$R_2 < R_4$	0,7832
2a	$R_3 \geq R_6$	$R_3 < R_6$	0,5084
2b	$R_4 \geq R_7$	$R_4 < R_7$	0,1280
3	$R_5 \geq R_{6,7}$	$R_5 < R_{6,7}$	0,1825
4	$R_{1,2,3,4} \geq R_{5,6,7}$	$R_{1,2,3,4} < R_{5,6,7}$	0,1390
5	$R_{13,4} \geq R_{6,7}$	$R_{3,4} < R_{6,7}$	0,3306
6	$R_{1,2,5} \geq R_{3,4,6,7}$	$R_{1,2,5} < R_{3,4,6,7}$	0,1186

8.5.3 VRAGENLIJST

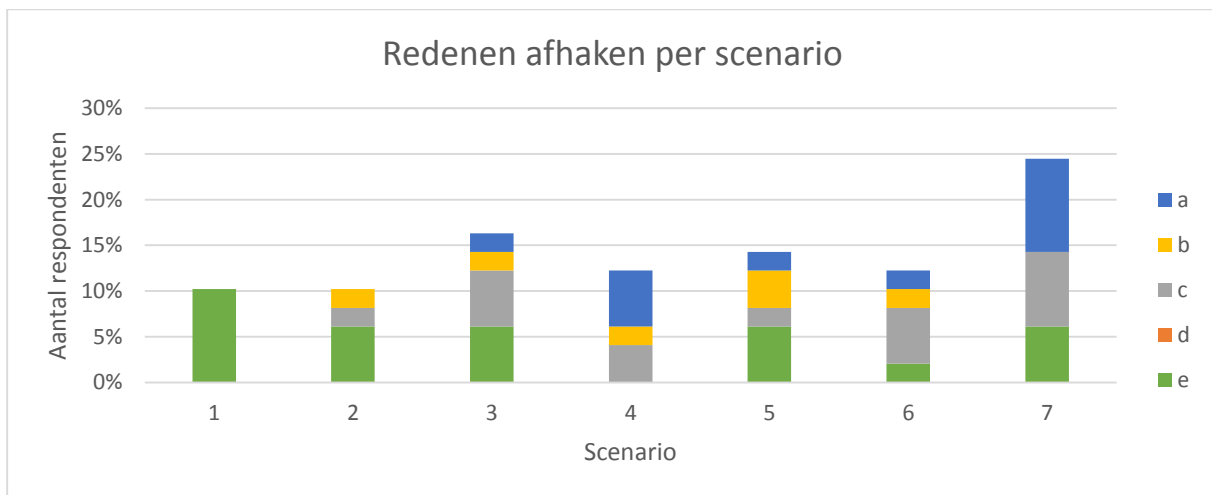
Na het doorlopen van de webshop en het al dan niet aanschaffen van een boek, wordt de deelnemer een pagina getoond met een vragenlijst. De inhoud van deze lijst is afhankelijk van hoe het voorgaande deel van de test is verlopen. Van de 174 respondenten hebben 146 het formulier ingevuld. Van deze 146 zijn er 91 respondenten geweest die eerder zijn afgehaakt. De overige 55 respondenten hebben het bestelproces doorlopen en een boek aangeschaft.

Van de vragenlijst zullen de belangrijkste resultaten nu worden behandeld. Dit betekent dat de reden van het afronden dan wel afhaken centraal staat.

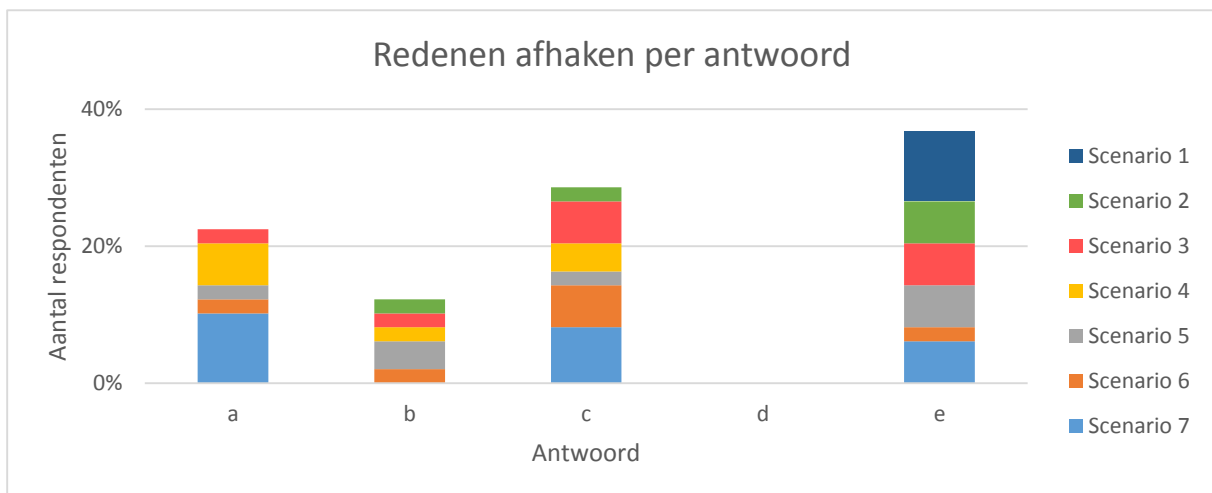
8.5.3.1 AFHAKERS

Figuur 8.6 en 8.7 geven een overzicht van de antwoorden op de vraag waarom de respondenten het boek niet hebben besteld en dus eerder zijn afgehaakt. De mogelijke antwoorden zijn:

- a) Webshop oogt niet betrouwbaar genoeg.
- b) Ik kon het boek niet vinden.
- c) Het duurde te lang voordat de pagina er was.
- d) De opdracht was onduidelijk.
- e) Anders, namelijk (invulveld).



Figuur 8.6 – Redenen voor respondenten om af te haken, gegroepeerd per scenario.



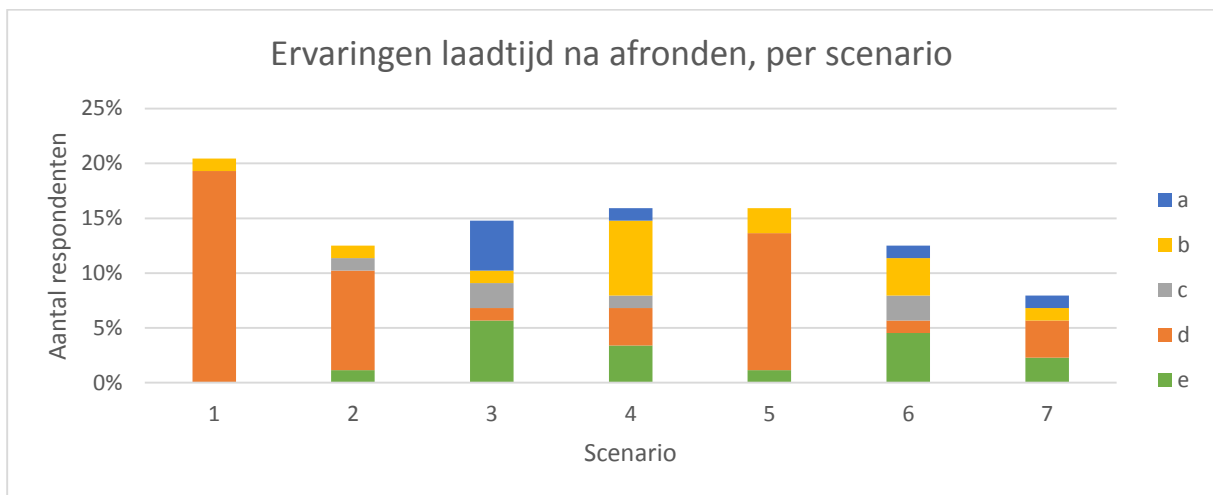
Figuur 8.7 – Redenen voor respondenten om af te haken, gegroepeerd per scenario.

Wanneer respondenten hebben gekozen voor antwoord e, bleek de meest opgegeven reden voor het afhaken tijdens het proces van inhoudelijke aard te zijn. Voorbeelden hiervan zijn dat men wenste ook met acceptgiro te kunnen betalen of dat men niet meerdere producten tegelijk in het winkelmandje kon bewaren.

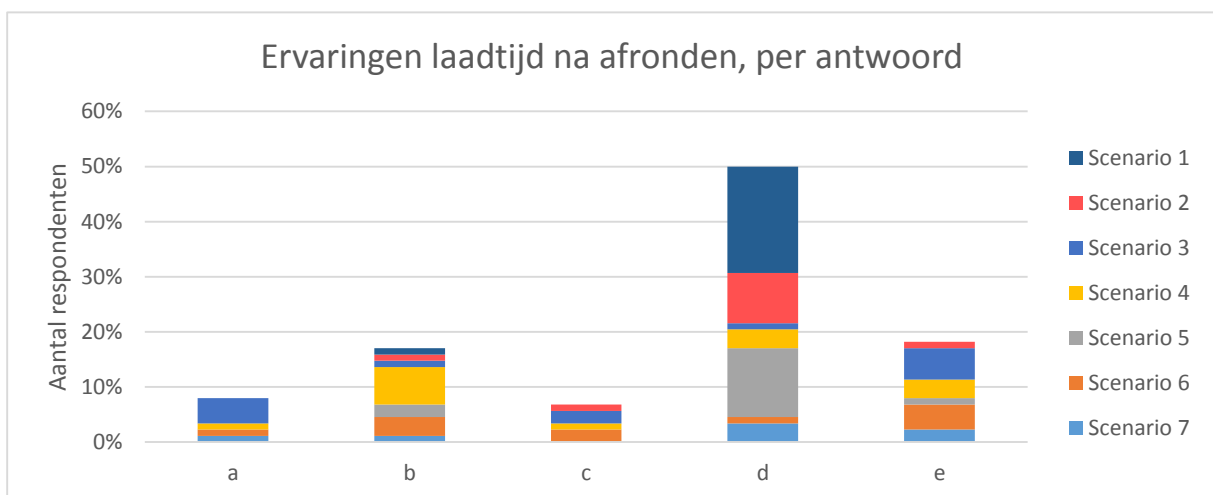
8.5.3.2 AFRONDERS

De grafiek in Figuur 8.8 en 8.9 geven een overzicht van de antwoorden op de vraag wat de deelnemers die een boek hebben aangeschaft vonden van de hogere laadtijd (indien ze die ervaren hadden). De mogelijke antwoorden zijn:

- Ik vind het niet erg, het is tenslotte een test.
- Ik dacht dat het met de test te maken had, dus daarom heb ik me er niets van aangetrokken.
- Geen probleem, normaal vind ik het ook niet erg wanneer ik iets langer moet wachten in een webshop.
- Trage pagina's? Daar heb ik niets van gemerkt.
- Anders, namelijk (invulveld).



Figuur 8.8 – Ervaringen laadtijd na afronden bestelling, gegroepeerd per scenario.



Figuur 8.9 – Ervaringen laadtijd na afronden bestelling, gegroepeerd per antwoord.

Opvallend is dat uit Figuur 8.9 blijkt dat de meest gekozen reactie op de vraag hoe de trage websites werden ervaren, de meeste respondenten geven aan dat ze dat niet hebben gemerkt.

8.5.3.3 INVULLING VAN DE BERICHTGEVING

Ook kregen alle respondenten de vraag voorgeschoteld welke boodschap hen zou overhalen toch op de site te blijven wanneer de pagina's traag zijn. Hierbij was alleen een eigen antwoord mogelijk. Het merendeel van de respondenten die hierop antwoord gaven zei een kortere melding op prijs te stellen met daarin excuses en een reden voor de traagheid (59%). Het op twee na meest genoemde antwoord is het concreet aangeven van de wachttijd (15%). Antwoorden die ook, maar in veel mindere mate, voorkwamen zijn onder andere het korting geven op producten en het tonen van iets wat afleidt van het wachten. Ook zijn er respondenten die aan hebben gegeven dat een melding geen invloed heeft op hun bezoek (17%).

8.6 CONCLUSIE

De acht nulhypothese van de test kunnen geen van allen verworpen worden. Dit betekent dat er niets kan worden gezegd over de gevolgen van de informatievoorziening ten tijde van performance-degradatie. In eerste instantie lijkt het hierbij niet uitgesloten dat dit is te wijten aan de grote groep deelnemers die is afgehaakt omdat ze de opdracht niet duidelijk vonden (49%) (Figuur 8.7).

Naast het niet kunnen verwerpen van de nulhypothese heeft de test andere informatie opgeleverd. Door de ingevulde formulieren is het alsnog mogelijk geweest op het oog interessante data te verzamelen. Zo blijkt dat het grootste deel van de respondenten van mening is dat berichtgeving tijdens performance-degradatie van nut kan zijn indien het de juiste informatie bevat.

9.1 CONCLUSIE

Het onderzoek bevat een hoofdcomponent, namelijk het ontwikkelen van de methodiek voor de MeasureWorks-tag en een kleiner component, het opzetten van de test uit Hoofdstuk 8.

Het ontwikkelen van de methodiek voor de MeasureWorks-tag heeft er toe geleid dat er twee waardes, de zogenaamde ORC_1 en ORC_2 zijn ontwikkeld om gedrag van website-bezoekers te kunnen monitoren. De eerste waarde neemt de websitenavigatie van bezoekers in ogenschouw, de tweede waarde geeft inzicht in de invloed van de website-performance op de conversie.

Het validatieproces van de methodiek van de tag is door het ontbreken van de juiste test-websites niet toereikend geweest. Dit heeft er toe geleid dat er een test ontwikkeld diende te worden om de gevolgen van berichtgeving tijdens performance-degradatie in kaart te brengen. De resultaten laten zien dat er te weinig bewijs is om de verschillende nulhypothese, die stellen dat informatievoorziening tijdens performance-degradatie geen invloed op de bezoeker heeft, te verwerpen. Een kritische noot hierbij is dat een groot deel van de respondenten achteraf heeft aangegeven de opdracht niet duidelijk te vinden. Dit zorgt ervoor dat er vraagtekens bij de betrouwbaarheid van de resultaten van de test kunnen worden gezet.

De enquête die na de test is ingevuld laat wél zien dat het merendeel van de respondenten informatie wil ontvangen wanneer er sprake is van performance-degradatie op een website. Dit ondersteunt de wetenschappelijke literatuur in Hoofdstuk 2. Hierdoor lijken met name de enquête na de test van essentieel belang te zijn.

Op basis van dit onderzoek lijkt het dan ook mogelijk om de performance-tolerantie van bezoekers van e-commerce-websites te verhogen ten tijde van performance-degradatie om zo de conversie te optimaliseren. Het inzetten van de MeasureWorks-tag maakt het mogelijk de website-performance te monitoren. De methodiek die aan de tag gekoppeld bepaald vervolgens op basis van de ORC_1 en ORC_2 wanneer het noodzakelijk is een bericht te sturen. Dit bericht zal dan door de MeasureWorks-tag naar de browser van de desbetreffende bezoeker worden gestuurd.

9.2 AANBEVELINGEN

Op het gebied van (het beïnvloeden van) performance-tolerantie is nog veel onbekend. Er is daarom een breed scala aan verdere onderzoeksmogelijkheden. Hieronder de belangrijkste mogelijkheden in het verlengde van dit onderzoek, weergegeven.

Het verbeteren van de huidige methodiek:

- De ORC_1 -waarde maakt een eerste onderscheid tussen converterende en niet-converterende bezoekers. Er is hier nog veel ruimte voor verbetering om de groepering te verfijnen. Een andere definitie van de ORC_1 -waarde kan hier misschien toe leiden.
- De momenten waarop de verschillende factoren van de methodiek worden gecreëerd alsmede de periode waarover deze factoren worden berekend, kunnen worden geoptimaliseerd. Omdat er in het huidige onderzoek te weinig testdata is verzameld, is dit nog niet gebeurd.
- Uitbreiden van de indeling van een websitebezoek in twee fasen: oriëntatiefase en bestelfase. Op basis van dit onderscheid kan er bijvoorbeeld rekening gehouden worden met de inhoud van een melding. Wanneer iemand zich al in het bestelproces bevindt kan het een meerwaarde opleveren een andere melding naar de bezoeker te sturen dan naar een bezoeker die zich nog in de oriëntatiefase bevindt. Bij de ontwikkeling van de methodiek is er al enigszins rekening gehouden met dit onderscheid. Er bleek helaas te weinig tijd over om dit verder uit te diepen.

Het onderzoeken van een derde ORC-waarde. Hierbij kan worden gedaan variabelen als:

- Operating system: is het operating system waarmee de bezoeker de website bezoekt van belang bij het verhogen van de tolerantie? Zou het bijvoorbeeld zo kunnen zijn dat gebruikers van Mac OS minder tolerant zijn voor een slechtere performance dan Windows-gebruikers.
- Leestijd: wat zegt de leestijd, de tijd waarop een pagina gelezen wordt, over de performance-tolerantie van een bezoeker.
- Object: als er naast Navigation Timing ook wordt beschikt over Resource Timing is het mogelijk op object-basis performance-waarden te monitoren. Op deze manier is het mogelijk concretere informatie over de performance te melden aan de bezoekers.

Daarnaast kan de inhoud van de berichtgeving aan bezoekers anders worden geformuleerd. Omdat de testresultaten niet hebben volstaan, is het interessant een nieuwe test op te zetten om er zo achter te komen wat in welke situatie de beste boodschap is om de performance-tolerantie te verhogen.

BRONNENLIJST

- 360°ecommerce (2014). *Wat is conversie optimalisatie?* Geraadpleegd op 24 juli 2014 via <http://www.360ecommerce.nl/wat-is-conversie-optimalisatie/>
- Bhatti, N., Bouchi, A. & Kuchinsku, A. (2000). *Integrating User-Perceived Quality into Web-Server Design*. *Computer Networks*, 33(1-6), 1-16.
- CanIUse (2014). *Can I use Navigation Timing API*. Geraadpleegd op 29 juni 2014 via <http://caniuse.com/nav-timing>.
- Dellaert, B.G.C & Kahn, B.E. (1999). *How Tolerable is Delay?: Consumers Evaluation of Internet Web Sites after Waiting*. *Journal of Interactive Marketing*, 13(1), 41-54.
- Forrester Consulting (2009). *eCommerce Web Site Performance Today*. Cambridge: Forrester Consulting.
- Harder, D. (Onbekend). *Numerical Analysis of Engineering*. Geraadpleegd op 24 juli 2014 via <https://ece.uwaterloo.ca/~dwharder/NumericalAnalysis/11Optimization/randombrute/>
- Kooiman, L. (2014). *Visualization of methodology MeasureWorks-tag*. Geraadpleegd op 27 juli 2014 via <http://beacon.mwtag.nl/lidewij/MeasureWorksTag.php?domain=wolfrijbewijsshop.nl>.
- Maister (1985). *The psychology of waiting lines*.
- Gomez (2010). *Why web performance matters: is your site driving customers away? Whitepaper*.
- MarketingExperiments (2009). *Beyond Landing Pages: Conversion rate optimization strategies*. MarketingExperiments.
- MarketingSherpa (2014). *Ecommerce Benchmark Study. Pre-release Executive Edition*. Magento.
- MathWorks (2014). *Least-Squares (Model Fitting) Algorithms*. Geraadpleegd op 13 juli 2014 via <http://www.mathworks.nl/help/optim/ug/least-squares-model-fitting-algorithms.html>. The MathWorks, Inc.
- Orion Web (2014). *Landing Page Optimization*. Geraadpleegd op 24 juli 2014 via <http://www.orionweb.net/milwaukee-internet-marketing/landing-page-optimization/>
- Que, A. (2013). *Gauss-Newton class*. Geraadpleegd op 14 februari 2014 via <http://gaussnewton.drque.net/>
- Radware (2014). *State of the Union: Ecommerce Page Speed & Web Performance*, Spring 2014. US: Radware.
- Rijksoverheid (2014). Geraadpleegd op 1 juli 2014 via <http://www.rijksoverheid.nl/onderwerpen/internet/bescherming-privacy-op-internet/cookiewet-regels-en-richtlijnen>.
- Seow, S.C. (2008). *Designing and engineering time*. New York: Addison Wesley.
- StatCounter (2014). *StatCounter Global Stats*. Geraadpleegd op 29 juni 2014 via <http://gs.statcounter.com/#desktop-browser-NL-monthly-201305-201405>.
- Velocity (2009). Geraadpleegd op 29-06-2014 via www.velocityconf.com. O'Reilly.
- W3C (2012). *Navigation Timing*. Geraadpleegd op 26 juni 2014 via <http://www.w3.org/TR/navigation-timing/#processing-modelBijlage>.
- Webics (2012). *What is a conversion funnel?* Geraadpleegd op 24 juli 2014 via <http://www.webics.com.au/blog/conversion-tracking/conversion-funnel/>
- Wolfram Alpha (2014). *Nonlinear Least Squares Fitting*. Geraadpleegd op 13 juli 2014 via <http://mathworld.wolfram.com/NonlinearLeastSquaresFitting.html>. Wolfram Research, Inc.
- Weisstein, E. (Onbekend). *Fisher's Exact Test*. Geraadpleegd op 24 juli 2013 via <http://mathworld.wolfram.com/FishersExactTest.html>.

A. PHP CODE

A.1 METHODIEK.PHP

```

<?php
function methodiek( $sid, $navigationstart, $page_label, $domain, $loadeventstart){
    $db_conn_methodiek = pg_connect('host=localhost dbname=tagdata user=lidewij password=_h0t_nHer3 connect_timeout=5') or die
("Kan geen verbinding met de database krijgen.");
    $inserttimestamp = $navigationstart;
    $loadtime = ($loadeventstart - $inserttimestamp);

    //convert unixts to pg datetime + tz
    $inserttimestamp_sec = round($inserttimestamp / 1000);
    $inserttimestamp = new DateTime("@$inserttimestamp_sec");
    date_timezone_set($inserttimestamp, timezone_open('Europe/Amsterdam'));
    $inserttimestamp = date_format($inserttimestamp, 'Y-m-d H:i:s T');
    $currentdate = date('Y-m-d H:i:s T', time());
    $firstlabelarray = explode('_', $page_label);
    $page_label = $firstlabelarray[0];

    if ( preg_match( "/^(https?:\V)?www\.(.+)$/" , $domain, $matches ) == 1){
        $domain = preg_replace('#^https?://www\.#', "", $matches[0]);
    } else {
        $domain = preg_replace('#^https?://#', "", $domain);
    }
    $domain = preg_replace('#/.*#', "", $domain);
    $domain = strtolower($domain);

    $queryTest = "SELECT * FROM tagdata.m_pagelabel_conversion WHERE domain = " . $domain . """;
    $result1 = pg_query($db_conn_methodiek, $queryTest) or die ( $queryTest . "<br />\n". pg_last_error() );
    if (pg_num_rows($result1) > 0 && $inserttimestamp < $currentdate ){

        ////////////////////////////////////////////////////////////////////*****DAADWERKELIJKE METHODIEK*****\//////////////////////////////////////////////////////////////////
        $insert_needed = true; //als er iets fout gaat met sid, dan notification == 10
        $orc1_exceeded = false;
        $orc2_exceeded = false;
        $notification = 0;
        $avg_orc1 = 0;
        $avg_orc2 = 0;

        //deze query haalt van gegeven $sid op: sessielengte, laatste pageview en max(notification_sent)
        $query = "SELECT COUNT(sid) AS session_length, MAX( inserttimestamp) AS last_insert, MAX(notification_sent) as notification FROM
tagdata.m_pageviews WHERE sid = " . $sid . """;
        $result = pg_query($db_conn_methodiek, $query) or die ( $query . "<br />\n". pg_last_error() );
        $row = pg_fetch_row($result);

        //als sid nieuw is;
        if( $row[0] == 0){
            $from_page_label = 'start';
            $avg_orc1_old = 0;
            $number_of_pageviews = 0;
            $notification_sent = 0;

        //als sid als bekend is;
        } else {
            $number_of_pageviews = $row[0];
            $last_insert = $row[1];
            $notification_sent = $row[2];

            if ($last_insert >= $inserttimestamp){
                $notification = 8; //timestamp hetzelfde
            }

            //zoekt meest recente pageview op
            $query = "SELECT page_label, orc1 FROM tagdata.m_pageviews WHERE inserttimestamp = " . $last_insert . " AND sid = " . $sid . """;
            $result = pg_query($db_conn_methodiek, $query) or die ( $query . "<br />\n". pg_last_error() );
            if (pg_num_rows($result) == 0){

```

```

    $notification = 2;
} else {
    //define page_label from last pageview
    $from_page_label = pg_fetch_result($result,0,0);
    $avg_orc1_old = pg_fetch_result($result,0,1);
    if ($avg_orc1_old < 0 || $avg_orc1_old > 1000){
        $notification = 3;
    }
}
}
if( $notification == 0){
    ////////////////////////////////////////////***** ORC1 *****//////////////////////////////////////////
    //probeert de orc1 op te halen vanuit de tabel orc1
    $query_orc1 = "SELECT transition_prob FROM tagdata.m_orc1 WHERE page_label_from = ". $from_page_label . " "
        . "AND page_label_to = ". $page_label . " AND domain = " . $domain . " ";
    $result_orc1 = (pg_query($db_conn_methodiek, $query_orc1)) or die ( $query_orc1 . "<br>". pg_last_error() );

    //kijkt of er daadwerkelijk data in de orc1-tabel zit
    if (pg_num_rows($result_orc1) == 0 ){

        //kijkt of er dan in de andere orc1 tabel wel data zit. Checkt eerst of deze tabel uberhaupt bestaat.
        $query_exists = "SELECT EXISTS(SELECT * FROM pg_tables WHERE schemaname='tagdata' AND tablename = 'm_orc1_copy')";
        $result_exists = (pg_query($db_conn_methodiek, $query_exists)) or die ( $query_exists . "<br />\n". pg_last_error() );
        $row = pg_fetch_row($result_exists);

        if ($row[0]== 't'){
            $query_orc1 = "SELECT transition_prob FROM tagdata.m_orc1_copy WHERE page_label_from = ". $from_page_label . " AND
page_label_to = ". $page_label . " AND domain = " . $domain . " ";
            $result_orc1 = (pg_query($db_conn_methodiek, $query_orc1)) or die ( $query_orc1 . "<br />\n". pg_last_error() );

            //Kijkt er of er data in de orc1_copy tabel zit (wanneer het origineel van orc1 gevuld wordt)
            if (pg_num_rows($result_orc1) == 0 ){
                $notification = 2;
            }
        } else {
            $avg_orc1 = 0;//insert in pageviews???
            //$notification = 33;
        }
    } else {
        //orc1-waarde gevonden in m_orc1
        $trans_prob = pg_fetch_result($result_orc1,0,0);
        $avg_orc1 = round(($trans_prob + $number_of_pageviews * $avg_orc1_old)/($number_of_pageviews + 1));

        $orc1_exceeded = 0;
        if( $avg_orc1 >= 500){
            $orc1_exceeded = true;
        }
    }
    ////////////////////////////////////////////***** ORC2 *****//////////////////////////////////////////
    //haalt de ORC2value.
    $query_orc2 = "SELECT AVG(orc2value)::integer FROM tagdata.m_orc2 WHERE domain = " . $domain . " ";
    $result_orc2 = (pg_query($db_conn_methodiek, $query_orc2)) or die ( $query_orc2 . "<br />\n". pg_last_error() );

    if (pg_num_rows($result_orc2) == 0 ){
        $notification = 2;
    } else {
        $avg_orc2 = pg_fetch_result ( $result_orc2 , 0 , 0 );
        if ($avg_orc2 < -1000){
            $orc2_exceeded = true;
        }
    }

    if ($orc2_exceeded && $orc1_exceeded ){
        $notification = 1;
    }
}

//////////////////////////////////////////***** ACTIES *****//////////////////////////////////////////
if ($notification == 0 || $notification == 1){

```

```

//voeg nieuwe pageview aan tabel pageviews toe
$queryInsert = "INSERT INTO tagdata.m_pageviews VALUES ('".$sid."', "" . $inserttimestamp .":TIMESTAMP," . $avg_orc1 . "," .
$page_label . "','";
    $notification . "," . $domain . "',' . $avg_orc2 . "," . $loadtime . "');";
    $insertPageview = pg_query($db_conn_methodiek, $queryInsert) or die ( $queryInsert . "<br />\n" . pg_last_error() );
} else {

    $queryInsert = "INSERT INTO tagdata.m_pageviews VALUES ('".$sid."', "" . $inserttimestamp .":TIMESTAMP,0," . $page_label . "','";
    .
    $notification . "," . $domain . "',' , 0," . $loadtime . "');";
    $insertPageview = pg_query($db_conn_methodiek, $queryInsert) or die ( $queryInsert . "<br />\n" . pg_last_error() );
}
}
}

```

A.2 ORC1.PHP

```

<?php
/*
/* Name: orc1.php
/* Date: 05/06/2013
/* Author: Lidewij Kooiman
/*
/* Vult tabel tagdata.m_orc1 met ORC1-waarde: q-overgangskansen
/* 1) maak kopie tagdata.m_orc1 --> tagdata.m_orc1-copy zodat orc1-data
/* beschikbaar blijft in methodiek
/* 2) leegt tagdata.m_orc1 en vult hem vervolgens weer met overgangskansen
/* Deze overgangskansen zijn gebaseerd op de navigatiewijze van de con-
/* verterende bezoeker. pxy = P(x --> y | x). Sommer over x: Som = 1.
/* qxy = som pxj >= 0.1 S=(j:pxj >= 0.1) als pxy >=0.1
/* qxy = som pxj < 0.1 S=(j:pxj <0.1) als pxy < 0.1
/*
/* Runt via crontab
/*****
require_once("tagDBlogin_methodiek.php");
$queryDropCopyOrc1Copy = "DROP TABLE IF EXISTS tagdata.m_orc1_copy";
pg_query($db_conn, $queryDropCopyOrc1Copy) or die ( $queryDropCopyOrc1Copy . "<br />\n" . pg_last_error() );

//kopieer originele orc1-tabel naar een copy: orc1_copy.
$query = "CREATE TABLE tagdata.m_orc1_copy AS SELECT * FROM tagdata.m_orc1";
$result = pg_query($db_conn, $query) or die ( $query . "<br />\n" . pg_last_error() );

//leegt originele orc1-tabel
$query = "TRUNCATE TABLE tagdata.m_orc1";
$result = pg_query($db_conn, $query) or die ( $query . "<br />\n" . pg_last_error() );

//voegt nieuwe waarden toe in orc1-tabel
$query =
" INSERT INTO tagdata.m_orc1 (
    SELECT CASE WHEN domein IS NULL
                THEN '-no domain defined-'
                ELSE CASE
                    WHEN domein = ''
                    THEN '-no domain defined-'
                    ELSE domein
                END
            END as domein,
        CASE
            WHEN vorig_pagelabel IS NULL
            THEN '-no pagelabel defined-'
        ELSE
            CASE
                WHEN vorig_pagelabel = ''
                THEN '- no pagelabel defined-'
                ELSE vorig_pagelabel
            END
            END as vorig_pagelabel,
        CASE
            WHEN pagelabel IS NULL
            THEN '-no pagelabel defined-'
            ELSE
                CASE
                    WHEN pagelabel = ''
                    THEN '- no pagelabel defined-'
                    ELSE pagelabel
                END
            END
        END as pagelabel,

```

```

ROUND(1000*(aantal_pl / (SUM(aantal_pl) OVER pl))::FLOAT)::INT AS fractie

FROM (
    SELECT
        domein,
        vorig_pagelabel,
        pagelabel,
        COUNT(vorig_pagelabel) AS aantal_pl

    FROM (
        SELECT
            CASE
                WHEN SUBSTRING( LOWER(url) FROM '.*://([^\/*]*)' ) LIKE 'www.%'
                THEN SUBSTRING( SUBSTRING( LOWER(url) FROM '.*://([^\/*]*)' ),5)
                ELSE SUBSTRING( LOWER(url) FROM '.*://([^\/*]*)' )
            END AS domein,
            ROW_NUMBER() OVER sessie AS rij_nummer,
            navigationStart,
            split_part( pagelabel, '_', 1) as pagelabel,
            CASE WHEN (ROW_NUMBER() OVER sessie)=1 THEN 'start'
            ELSE
                CASE
                    WHEN (LAG(split_part( pagelabel, '_', 1),1) OVER sessie) IS
NULL
                    THEN '-no pagelabel defined-'
                    ELSE LAG(split_part( pagelabel, '_', 1),1) OVER sessie
                END
            END AS vorig_pagelabel,
            z.sid
        FROM tagdata." . ""tagData" . " AS z

        JOIN (
            SELECT distinct sid
        FROM tagdata." . ""tagData" . "
        WHERE split_part( pagelabel, '_', 1) IN (SELECT pagelabel FROM tagdata.m_pagelabel_conversion)
        AND navigationStart >= CURRENT_TIMESTAMP - INTERVAL '7 DAY') as ids
        ON ids.sid = z.sid

        WHERE navigationStart >= CURRENT_TIMESTAMP - INTERVAL '7 DAY'
        WINDOW sessie AS (PARTITION BY z.sid ORDER BY navigationStart)
    ) AS basis
WHERE domein IN (SELECT domain FROM tagdata.m_pagelabel_conversion)
GROUP BY basis.domein, basis.pagelabel, basis.vorig_pagelabel
ORDER BY basis.domein, vorig_pagelabel, pagelabel

) AS tweede

WHERE domein != '-no domain defined-' AND
vorig_pagelabel != '-no pagelabel defined-' AND
aantal_pl > 2 AND
pagelabel != '-no pagelabel defined-'
WINDOW pl AS (PARTITION BY domein, vorig_pagelabel)
ORDER BY domein, vorig_pagelabel
);

UPDATE tagdata.m_orc1
SET transition_prob = s.som
FROM (SELECT domain as joindomain , page_label_from as labelfrom, sum(transition_prob) as som
FROM tagdata.m_orc1
WHERE transition_prob >= 100
GROUP BY joindomain, page_label_from) as s
WHERE page_label_from = labelfrom AND
domain = joindomain
AND transition_prob >= 100;

UPDATE tagdata.m_orc1
SET transition_prob = s.som
FROM (SELECT domain as joindomain, page_label_from as labelfrom, sum(transition_prob) as som
FROM tagdata.m_orc1
WHERE transition_prob < 100
GROUP BY joindomain, page_label_from) as s
WHERE page_label_from = labelfrom AND
joindomain = domain
AND transition_prob < 100;";

```

```

$result = pg_query($db_conn, $query) or die ( $query . "<br />\n". pg_last_error() );
//Verwijderd kopie van orc1-tabel; de data kan nu weer uit de originele tabel gehaald worden.
$query = "DROP TABLE tagdata.m_orc1_copy";
$result = pg_query($db_conn, $query) or die ( $query . "<br />\n". pg_last_error() );
pg_close($db_conn);
echo "end ok";
?>

```

A.3 ORC2.PHP

```

<?php
require_once("tagDBlogin_methodiek.php");
require_once("orc2Functions.php");
require_once("GaussNewton/ExponentialRegression.php");
require_once("GaussNewton/GaussNewtonRegression.php");

/* De tabel die hierbij wordt ge-insert in 'm_sessionlength' geeft per:
- domain
- pagelabel
- sessielengte
een overzicht van de gemiddelde laadtijd. Bijvoorbeeld:
- domain: www.measureworks.nl
- pagelabel: contact
- sessielengte: lengte van de sessie waarbinnen deze pageview valt

1) Selecteert domain, pagelabel, loadeventstart, navigationstart en sessielengte per pageviews
2) Vervolgens worden er een domain en pagelabel gedefinieerd als deze niet zijn. De sessielengte
wordt doorgegeven en de laadtijd wordt berekend door loadeventstart - navigationstart (in ms)
3) Tot slot wordt per domain, pagelabel en sessielengte de gemiddelde laadtijd berekend.*/

echo date('m/d/Y h:i:s a', time()) . "<br>";
$queryCreateCopySessionlength = "CREATE TABLE IF NOT EXISTS tagdata.m_sessionlength_copy (LIKE tagdata.m_sessionlength);";
INSERT INTO tagdata.m_sessionlength_copy (SELECT * FROM tagdata.m_sessionlength);

$queryTruncateSessionlength = "TRUNCATE TABLE tagdata.m_sessionlength;";
$queryInsertIntoSessionlength = "
INSERT INTO tagdata.m_sessionlength (
SELECT
s.domain as domain,
CASE WHEN s.domain = '-no domain defined-'
THEN '- no pagelabel defined-'
ELSE s.pagelabel
END as pagelabel,
s.sessionlength,
count(*)::integer as aantal_pageviews,
ROUND(AVG(s.loadtime))

FROM
(SELECT
t.sid,CASE WHEN t.domain IS NULL
THEN '-no domain defined-'
ELSE CASE WHEN t.domain = ''
THEN '-no domain defined-'
ELSE t.domain
END
END,
CASE WHEN t.pagelabel IS NULL
THEN '-no pagelabel defined-'
ELSE CASE WHEN t.pagelabel = ''
THEN '- no pagelabel defined-'
ELSE t.pagelabel
END
END,
t.sessionlength,
(EXTRACT(epoch FROM (t.loadeventstart - t.navigationstart))*1000)::integer AS loadtime

FROM(
SELECT
sid,
CASE WHEN SUBSTRING( LOWER(url) FROM '.*://([/^\w]*)' ) LIKE 'www.%'

```



```

        THEN SUBSTRING( SUBSTRING( LOWER(url) FROM '.*://([^\/*]*)' ),5)
        ELSE SUBSTRING( LOWER(url) FROM '.*://([^\/*]*)' )
    END AS domain,
    split_part( pagelabel, '_', 1) AS pagelabel,
    loadeventstart,
    navigationstart,
    count(sid) OVER (PARTITION BY sid)::smallint AS sessionlength
FROM tagdata." . ""tagData"" . "
WHERE navigationStart >= CURRENT_TIMESTAMP - INTERVAL '26 HOUR'

) as t
WHERE t.navigationStart >= CURRENT_TIMESTAMP - INTERVAL '24 HOUR' ) as s

WHERE s.loadtime < 3600000 AND
    domain != '-no domain defined-' AND
    pagelabel != '-no pagelabel defined-'
GROUP BY domain, pagelabel, sessionlength
ORDER BY domain, pagelabel, sessionlength, aantal_pageviews
);";

$queryDropCopySessionlength = "DROP TABLE IF EXISTS tagdata.m_sessionlength_copy;";

pg_query($db_conn, $queryCreateCopySessionlength) or die ( $queryCreateCopySessionlength . "<br />\n". pg_last_error() );
pg_query($db_conn, $queryTruncateSessionlength) or die ( $queryTruncateSessionlength . "<br />\n". pg_last_error() );
pg_query($db_conn, $queryInsertIntoSessionlength) or die ( $queryInsertIntoSessionlength . "<br />\n". pg_last_error() );
pg_query($db_conn, $queryDropCopySessionlength) or die ( $queryDropCopySessionlength . "<br />\n". pg_last_error() );

$queryTruncateORC2 = "TRUNCATE TABLE tagdata.m_orc2";
$data = pg_query($db_conn, $queryTruncateORC2) or die ( $queryTruncateORC2 . "<br />\n". pg_last_error() );
$queryGetDomainsAndPagelabels = "SELECT domain, pagelabel, count(*) as AANTAL FROM tagdata." . ""m_sessionlength"" . " WHERE
sessionlength <=20 GROUP BY domain, pagelabel ORDER BY domain, pagelabel";
$result = pg_query($db_conn, $queryGetDomainsAndPagelabels) or die ( $queryGetDomainsAndPagelabels . "<br />\n". pg_last_error() );
$aantalGrafieken = pg_num_rows( $result );

if ( $aantalGrafieken > 0 ) {
    //per domain en pagelabel wordt de data uit m_sessionlength gehaald
    //hierbij worden er arrays gecreeërd
    // $x = sessionlength, $y = loadtime, $w = frequency/total
    $numberOfGraphs = 0;

    //itereert door domain en pagelabel
    while ( $obs = pg_fetch_array( $result ) ) {
        $domain = $obs['domain'];
        $pagelabel = $obs['pagelabel'];

        $queryGetInputForOrc2 = "
        SELECT
            *,
            SUM(frequency) OVER (PARTITION BY domain, pagelabel) AS total
        FROM tagdata." . ""m_sessionlength"" . "
        WHERE domain = '" . $domain . "' AND
            pagelabel = '" . $pagelabel . "' AND
            sessionlength <= 20
        ORDER BY sessionlength ";

        $i = 0;
        $x = array();
        $y = array();
        $w = array();
        $frequency = array();
        $data = pg_query($db_conn, $queryGetInputForOrc2) or die ( $queryGetInputForOrc2 . "<br />\n". pg_last_error() );

        //doorloopt per domain en pagelabel de gemiddelde laadtijd/sessielengte
        while ( $obs = pg_fetch_array( $data ) ) {
            $orc2 = 0;
            $x[$i] = $obs['sessionlength'];
            $y[$i] = $obs['loadtime'];
            $w[$i] = $obs['frequency']/$obs['total'];
            $frequency[$i] = $obs['frequency'];
        }
    }
}

```

```

        $totalFrequency = $obs['total'];
        $i++;
    }
    $ydata = graphValues($x, $y);
    $wdata = graphValues($x, $w);
    $frequencydata = graphValues($x, $frequency);
    $methodiek = false;

    //zorg dat de methodiek alleen bij 10 of meer waarnemingen en min 3 verschillende sessielengtes wordt berekend.
    if( $totalFrequency > 10 && count($x) > 3){
        $schattingC = $y[count($x)-1] ;
        $schattingA = (($y[0] - $schattingC) * exp(1));
        $startcoefficients = array( $schattingA, -0.75, $schattingC);
        $startvalues = calculateExpFunction($x, $startcoefficients);
        $error_start = calculateSquareError($w, $y, $startvalues);
        $startdata = graphValues($x, $startvalues);
        $bestFit_error = $error_start;
        $bestFit_coefficients = $startcoefficients;
        $bestFit_data = $startdata;
        $numberOfCalls= 0;

        while ( $numberOfCalls < 200){ //4000? --> moet veel groter getal worden
            $coefficients = test( $x, $y, $w, $startcoefficients );
            $data = calculateExpFunction($x, $coefficients);
            $error_fit = calculateSquareError($w, $y, $data );
            if ($error_fit < $bestFit_error){
                $bestFit_coefficients = $coefficients;
                $bestFit_error = $error_fit;
                $bestFit_data = $data;
            }
            $numberOfCalls += 1;
        }
        $regression = new ExponentialRegression();
        $coefficients = $regression->refineCoefficients( $x, $y, $w, $bestFit_coefficients );
        $data = calculateExpFunction($x, $coefficients);
        $error_fit = calculateSquareError($w, $y, $data );
        $factor = 0.1;
        $Infactor = 2.30258509299405000;
        if ($bestFit_error < 100 && $bestFit_coefficients[0] >= 0 && $bestFit_coefficients[1] < 0 && $bestFit_coefficients[2] > 0 ){
            $orc2 = round(($bestFit_coefficients[0] * $bestFit_coefficients[1] * (1- $factor))/($bestFit_coefficients[2]*$Infactor),1);
            if( $orc2 < 0 && $orc2 > -100000){
                $fitdata = graphValues($x, $bestFit_data);
                $methodiek = true;
            }
        }
    }

    if( $methodiek == false){
        $bestFit_coefficients = array( 0,0,0);
        $orc2 = 0;
    }

    $part = 0;
    $queryGetPart = "SELECT part FROM tagdata.m_pagelabels WHERE page_label = " . $pagelabel . """;
    $dataGetPart = pg_query($db_conn, $queryGetPart) or die ( $queryGetPart . "<br />\n". pg_last_error() );
    if(pg_num_rows($dataGetPart) == 0){
        $part = 0;
    } else {
        $part = pg_fetch_result ( $dataGetPart , 0 , 0 );
    }
    $queryInsertORC2 = "INSERT INTO tagdata.m_orc2 VALUES(" . $domain . "," . $pagelabel . "," . $orc2 . "," .
        $part . "," . $bestFit_coefficients[0] . "," . $bestFit_coefficients[1] . "," . $bestFit_coefficients[2] . ")";
    pg_query($db_conn, $queryInsertORC2) or die ( $queryInsertORC2 . "<br />\n". pg_last_error() );
    $numberOfGraphs++;
}
}

$queryAddOrc2History = "
INSERT INTO tagdata.m_orc2_history(
SELECT

```

```

        CURRENT_TIMESTAMP,
        domain,
        page_label,
        orc2value,
        part
    FROM tagdata.m_orc2
);
pg_query($db_conn, $queryAddOrc2History) or die ( $queryAddOrc2History . "<br />\n". pg_last_error() );
pg_close($db_conn);
echo "end ok\n";

```

A.4 ORC2FUNCTIONS.PHP

```

<?php
//loop met iteraties voor beste coëfficiënten adhv brute force random search
function test( $x, $y, $w, $coefficients ){
    $count = 0;
    $beginCoefficients = $coefficients;
    $fitvalues;
    $errorfit_sqrt = 0;
    $whileloop = 100;

    for ( $index = 0; $index < count($x) ; ++$index ){
        $fitvalues[$index] = ROUND($coefficients[ 0 ] * exp( $coefficients[ 1 ] * ($x[$index] ) ) + $coefficients[ 2 ]);
        $errorfit_sqrt += ($w[$index])*( $fitvalues[$index] - $y[$index] )*( $w[$index] )*( $fitvalues[$index] - $y[$index] );
    }
    $oldfitererror_sqrt = $errorfit_sqrt;
    $starterror = $oldfitererror_sqrt;
    $sameerror = 0;

    while ( $count <= $whileloop ){
        if ( $errorfit_sqrt <= $oldfitererror_sqrt && $count < $whileloop ) {
            $oldCoefficients = $coefficients;
            $oldfitererror_sqrt = $errorfit_sqrt;
            $coefficients[0] = $coefficients[0] + rand(-$coefficients[0], $coefficients[0]);
            $coefficients[1] = $coefficients[1] + rand(-100,-1)/100;
            $coefficients[2] = $coefficients[2] + rand(-min(150, $coefficients[2]), +150);
        } else {
            $sameerror += 1;
        }
    }
    if ( $sameerror > 2 ){

        if ( is_nan($coefficients[0]) || is_nan($coefficients[1]) || is_nan($coefficients[2]) || $errorfit_sqrt > 100000 ){
            $coefficients[0] = $oldCoefficients[0] + rand(-$oldCoefficients[0], $oldCoefficients[0]);
            $coefficients[1] = rand(-1000,-1)/100;
            $coefficients[2] = $oldCoefficients[2] + rand(-min(300, $oldCoefficients[2]), +300);
        }
    }
    $errorfit_sqrt = 0;
    $fitvalues;
    for ( $index = 0; $index < count($x) ; ++$index ){
        $fitvalues[$index] = ROUND($coefficients[ 0 ] * exp( $coefficients[ 1 ] * ($x[$index] ) ) + $coefficients[ 2 ]);
        $errorfit_sqrt += ($w[$index])*( $fitvalues[$index] - $y[$index] )*( $w[$index] )*( $fitvalues[$index] - $y[$index] );
    }
    if( $count == $whileloop ){
        if( $errorfit_sqrt > $oldfitererror_sqrt ){
            $coefficients = $oldCoefficients;
        }
    }
    $count++;
}

return $coefficients;
}

function graphValues( $x, $y ){
    $ydata = "[";
    for( $j = 0; $j < count($x); $j++ ) {
        if ( $j < count( $x ) - 1 ){
            $ydata .= "[" . $x[$j] . " , " . $y[$j] . " ] , " ;
        } else {
            $ydata .= "[" . $x[$j] . " , " . $y[$j] . " ] " ;
        }
    }
}

```

```

    }
    }
    $ydata .= "];";
    return $ydata;
}

function calculateExpFunction($x, $coefficients){
//berekent y wanneer a,b,c en x gegeven zijn.
    $fitvalues = array();
    for ( $index = 0; $index < count($x); ++$index ){
        $fitvalues[ $index ] = ROUND($coefficients[ 0 ] * exp( $coefficients[ 1 ] * $x[$index] ) + $coefficients[ 2 ]);
    }
    return $fitvalues;
}

function calculateSquareError($w, $y, $fit){
//$y = daadwerkelijke pageviews
//$fit = gefitte punten (advh exp functie)
    $errorfit = 0;
    for ( $index = 0; $index < count($y); ++$index ){
        $errorfit += ($w[$index]*($fit[$index] - $y[$index])*$w[$index]*($fit[$index] - $y[$index]));
    }
    return ($errorfit/(array_sum($y)/count($y)));
}

```

A.4 EXPONENTIALREGRESSION.PHP

```

<?php
/*****
/* Name : ExponentialRegression.php */
/* Uses : Implementation of several forms for exponential regression. */
/* Date : July 14, 2014 */
/* Author: Lidewij Kooiman */
/* (C) Copyright 2013 by Andrew Que */
/* http://www.DrQue.net/ */
*****/

require_once("GaussNewtonRegression.php");
/**
 * Various exponential implementation of the Gauss-Newton algorithm.
 *
 * @package GaussNewton
 * @author Andrew Que ({@link http://www.DrQue.net/})
 * @copyright Copyright (c) 2013, Andrew Que
 * @license http://opensource.org/licenses/gpl-license.php GNU Public License
 */

class ExponentialRegression extends GaussNewtonRegression{
/**
 *  $f(x) = a * \exp(b * x) + c$ 
 * Regression for the exponential function with an offset.
 *
 * @package GaussNewtonRegression
 * @subpackage ExponentialRegression
 */

protected function partialDifferential( $x, $coefficientIndex, $coefficients ){
/**
 * Partial differential.
 *
 * @param float $x Value of x to supply to partial-differential function.
 * @param integer $coefficientIndex Which coefficient to be used in the
 * partial-differential function.
 * @param array $coefficients Values of the coefficients to be used.
 * @return float p( x ) for the supplied input.
 */
    assert( $coefficientIndex < 3 );
    $result = 0;
    switch ( $coefficientIndex ){

```

```

    case 0:{
        $result = exp( $coefficients[ 1 ] * $x );
        break;
    }

    case 1:{
        $result = $coefficients[ 0 ] * $x * exp( $coefficients[ 1 ] * $x );
        break;
    }

    case 2:{
        $result = 1;
        break;
    }

    default:{
        assert( false );
        break;
    }
}
return $result;
}

protected function getNumberOfCoefficients() {
/**
 * Number of coefficients.
 *
 * @return integer The number of coefficients for the equation being used.
 */
return 3;
}

public function getFunction( $x, $coefficients ){
/**
 * Evaluate function at x.
 *
 * Return f ( x ) for a given set of coefficients.
 *
 * @param float $x Real value given to equation.
 * @param array $coefficients Coefficients used in calculation.
 * @return float Value result of equation.
 */
return $coefficients[ 0 ] * exp( $coefficients[ 1 ] * $x ) + $coefficients[ 2 ];
}
}

```

A.5 GAUSSNEWTONREGRESSION.PHP → GEWICHTEN TOEGEVOEGD AAN ORIGINEEL

```

<?php
/*****
/* Name : GaussNewtonRegression.php */
/* Uses : Class for preforming Gauss-Newton regression. */
/* Date : 05/05/2013 */
/* Author: Andrew Que <http://www.DrQue.net/> */
/* Revisions: */
/* 1.0 - 05/05/2013 - QUE - Creation. */
/* Credits: */
/* Prof. C. Balaji of the Indian Institute of Technology Madras. */
/* http://mech.iitm.ac.in/Faculty/CB/home.php */
/* References: */
/* (1)...http://en.wikipedia.org/wiki/Matrix_multiplication */
/* (2)...http://en.wikipedia.org/wiki/Gauss%E2%80%93Newton_algorithm */
/* (3)...http://en.wikipedia.org/wiki/Gaussian_elimination */
/* (4)...http://en.wikipedia.org/wiki/Jacobian_matrix_and_determinant */
/* (5)...http://en.wikipedia.org/wiki/Coefficient_of_determination */
/* License: */
/* This program is free software: you can redistribute it and/or modify */
/* it under the terms of the GNU General Public License as published by */
/* the Free Software Foundation, either version 3 of the License, or */
/* (at your option) any later version. */

```

```

/*
/* This program is distributed in the hope that it will be useful,
/* but WITHOUT ANY WARRANTY; without even the implied warranty of
/* MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
/* GNU General Public License for more details.
/*
/* You should have received a copy of the GNU General Public License
/* along with this program. If not, see <http://www.gnu.org/licenses/>.
/*
/* (C) Copyright 2013 by Andrew Que
/* http://www.DrQue.net/
/*****

/**
 * Gauss-Newton non-linear regression.
 *
 * Used for calculating least-square regression coefficients. Useful for
 * linear and non-linear curve fitting.
 *
 * @package GaussNewton
 * @author Andrew Que (http://www.DrQue.net/)
 * @copyright Copyright (c) 2013, Andrew Que
 * @license http://opensource.org/licenses/gpl-license.php GNU Public License
 */

/**
 * Gauss-Newton non-linear regression.
 * This class is useful for computing the coefficients for some non-linear
 * equation using the Gauss-Newton algorithm.
 *
 * Algorithm (see 2):
 *  $b_{n+1} = b_n + (J^T J)^{-1} J r(b_n)$ 
 * Where b is the current guess for the coefficients, n is the current guess
 * J is the Jacobian, and r is the residual (error).
 *
 * This is an abstract class and can not be used by itself. It requires the
 * the function and function partial differentials to be define.
 *
 * @package GaussNewtonRegression
 * @abstract
 */
abstract class GaussNewtonRegression
{

    // Implementation notes:
    // All matrices in this class are assumed to be a two-dimensional array
    // indexed by row and then column. Matricies are only used internally, and
    // all public input/output is done using one-dimensional arrays.
    // Matrix inversion is not implemented because the multiplying the a matrix
    // by the an other matrix's inverse is the same as solving a system of
    // equations. Solving this system directly saves the multiplication step.
    //

    /**
     * Compute partial-differential of x.
     *
     * Compute the partial-differential with respect to a given coefficient for
     * a given value of x.
     *  $p(x) = df(x) / dc_n$ 
     * Example:
     * <code>
     * If  $f(x) = c_0 e^{c_1 x}$ , then:
     *  $df(x) / dc_0 = e^{c_1 x}$ 
     *  $df(x) / dc_1 = c_0 x e^{c_1 x}$ 
     * Each function must be defined.
     * </code>
     * @param float $x Value of x to supply to partial-differential function.
     * @param integer $coefficientIndex Which coefficient to be used in the
     * partial-differential function.
     * @param array $coefficients Values of the coefficients to be used.

```

```

* @return float p( x ) for the supplied input.
*/
abstract protected function partialDifferential
(
    $x,
    $coefficientIndex,
    $coefficients
);
abstract public function getFunction( $x, $coefficients );
/**Return the transpose of a matrix.
* @param array $matrix Any matrix.
* @return array A matrix representing transpose. */
private function transpose( $matrix )
{
    $result = array();
    for ( $row = 0; $row < count( $matrix[ 0 ] ); ++$row )
    {
        $result[ $row ] = array();
        for ( $column = 0; $column < count( $matrix ); ++$column )
            $result[ $row ][ $column ] = $matrix[ $column ][ $row ];
    }
    return $result;
}
/**
* Multiply two matrices (see 1).
* [ A ] = [ B ][ C ]
* @param array $matrix - Matrix [ B ].
* @param array $multiplicand - Matrix [ C ].
* @return array Matrix [ C ].
*/
private function multiply( $matrix, $multiplicand )
{
    $result = array();
    for ( $row = 0; $row < count( $matrix ); ++$row )
    {
        $result[ $row ] = array();
        for ( $column = 0; $column < count( $multiplicand[ 0 ] ); ++$column )
        {
            $result[ $row ][ $column ] = 0;
            for ( $index = 0; $index < count( $matrix[ $row ] ); ++$index )
                $result[ $row ][ $column ] +=
                    $matrix[ $row ][ $index ] * $multiplicand[ $index ][ $column ];
        }
    }
    return $result;
}
/**
* Multiply matrix with vector (see 1).
* [ A ] = [ B ] * c
* @param array $matrix - Matrix [ B ].
* @param array $vector.
* @return array Matrix [ C ].
*/
private function multiplyWithVector( $matrix, $vector )
{
    $result = array();
    for ( $row = 0; $row < count( $matrix ); ++$row )
    {
        $result[ $row ] = 0;
        for ( $column = 0; $column < count( $vector ); ++$column ) {
            $result[ $row ] += $matrix[ $row ][ $column ] * $vector[ $column ];
        }
    }
    return $result;
}
/** * Get a vector of how much error is between the calculated value, and the
* true value.
*
* @param float $x Array of x-coordinates.

```

```

* @param float $y Array of known y-coordinates that correspond to $x.
* @param array $coefficients Current coefficients to check for error.
* @return vector Matrix of error at each point.
*/
private function getErrorMatrix( $x, $y, $coefficients )
{
    assert( count( $x ) == count( $y ) );

    $result = array();
    for ( $row = 0; $row < count( $x ); ++$row )
        $result[ $row ] = $y[ $row ] - $this->getFunction( $x[ $row ], $coefficients);

    return $result;
}
/*
* Solve a system of equations in matrix form.
*
* Done by concatenating answer matrix to the left, doing Gaussian
* elimination (see 3), and returning the last column.
* [ M ][ C ] = [ A ] ==> [ C ] = [ M ]^-1 [ A ]
*
* @param array $matrix - Square matrix [ M ].
* @param array $answers - Single column matrix [ A ].
* @output array Single column matrix [ C ].
*/
private function solve( $matrix, $answers )
{
    $degree = count( $matrix );
    $order = array();
    //-----
    // Add the answers to the matrix.
    //-----
    $isDone = array();
    for ( $row = 0; $row < $degree; ++$row )
    {
        $matrix[ $row ][ $degree ] = $answers[ $row ];
        $isDone[ $row ] = false;
    }

    //-----
    // We now row-reduce the matrix to obtain the coefficients for the
    // polynomial.
    //-----

    // This loop will result in an upper-triangle matrix with the
    // diagonals all 1--the first part of row-reduction--using 2
    // elementary row operations: multiplying a row by a scalar, and
    // subtracting a row by a multiple of an other row.
    // NOTE: This loop can be done out-of-order. That is, the first
    // row may not begin with the first term. Order is tracked in the
    // "Order" array.
    $order = array();
    for ( $column = 0; $column < $degree; ++$column )
    {
        // Find a row to work with.
        // A row that has a term in this column, and hasn't yet been
        // reduced.
        $activeRow = 0;
        while ( ( 0 == $matrix[ $activeRow ][ $column ] ) || ( $isDone[ $activeRow ] ) ) && ( $activeRow < $degree )
        {
            ++$activeRow;
        }

        // Do we have a term in this row?
        if ( $activeRow < $degree )
        {
            // Remeber the order.
            $order[ $column ] = $activeRow;

            // Normilize row--results in the first term being 1.
            $firstTerm = $matrix[ $activeRow ][ $column ];

```



```

for ( $subColumn = $column; $subColumn <= $degree; ++$subColumn )
    $matrix[ $activeRow ][ $subColumn ] /= $firstTerm;

// This row is finished.
$isDone[ $activeRow ] = true;

// Subtract the active row from all rows that are not finished.
for ( $row = 0; $row < $degree; ++$row )
    if ( ( ! $isDone[ $row ] )
        && ( 0 != $matrix[ $row ][ $column ] ) )
    {
        // Get first term in row.
        $firstTerm = $matrix[ $row ][ $column ];
        for ( $subColumn = $column; $subColumn <= $degree; ++$subColumn )
            $matrix[ $row ][ $subColumn ] -=
                $firstTerm * $matrix[ $activeRow ][ $subColumn ];
    }
}

// Reset done.
for ( $row = 0; $row < $degree; ++$row )
    $isDone[ $row ] = false;

// Storage for the resulting coefficients.
$coefficients = array();

// Back-substitution.
// This will solve the matrix completely, resulting in the identity
// matrix in the x-locations, and the coefficients in the last column.
// | 1 0 0 ... 0 c0 |
// | 0 1 0 ... 0 c1 |
// | . . . . . |
// | . . . . . |
// | 0 0 0 ... 1 cn |
for ( $column = ($degree - 1); $column >= 0; --$column )
{
    // The active row is based on order.
    $activeRow = $order[ $column ];

    // The active row is now finished.
    $isDone[ $activeRow ] = true;

    // For all rows not finished...
    for ( $row = 0; $row < $degree; ++$row )
        if ( ! $isDone[ $row ] )
        {
            $firstTerm = $matrix[ $row ][ $column ];

            // Back substitution.
            for ( $subColumn = $column; $subColumn <= $degree; ++$subColumn )
                $matrix[ $row ][ $subColumn ] -=
                    $firstTerm * $matrix[ $activeRow ][ $subColumn ];
        }

    // Save this coefficients for the return.
    $coefficients[ $column ] = $matrix[ $activeRow ][ $degree ];
}

return $coefficients;
}

/**
 * Refine coefficients one around.
 *
 * Run an iteration of the Gauss-Newton algorithm. Run this as many times
 * as needed to refine coefficients to their best values.
 *
 * @param array $x Array of x data points.
 * @param array $y Array of y data points that correspond to x.
 * @param array $coefficients Initial or previous guess at coefficients.

```

```

* @return array An array that is the new guess at coefficients.
*/
public function refineCoefficients( $x, $y, $w, $coefficients )
{
    // Number of coefficients.
    $numberOfCoefficients = 3;
    $numberOfDatapoints = count($x);

    // Compute Jacobian matrix (see 4).
    // The Jacobian matrix consists of the partial-differentials.
    // [ d f(x_0) / d c_0  d f(x_0) / d c_1  ...  d f(x_0) / d c_m ]
    // [ d f(x_1) / d c_0  d f(x_1) / d c_1  ...  d f(x_1) / d c_m ]
    // [ . . . . . ]
    // [ . . . . . ]
    // [ . . . . . ]
    // [ d f(x_n) / d c_0  d f(x_n) / d c_1  ...  d f(x_n) / d c_m ]
    //
    // Where f() is the non-linear function, d f() / d c is the partial-
    // differential with respect to c. n is the number of rows in x, and m
    // is the number of coefficients in the equation.
    $jacobianMatrix = array();
    $weights = array();

    for ( $row = 0; $row < count( $x ); ++$row )
    {
        $jacobianMatrix[ $row ] = array();
        $weights[ $row ] = array();
        for ( $column = 0; $column < $numberOfCoefficients; ++$column ){
            $jacobianMatrix[ $row ][ $column ] = $this->partialDifferential( $x[ $row ], $column, $coefficients );
        }

        for ( $column = 0; $column < $numberOfDatapoints; ++$column ){
            if ( $row == $column ){
                $weights[ $row ][ $column ] = $w[ $row ];
            } else {
                $weights[ $row ][ $column ] = 0;
            }
        }
    }

    // Compute the transpose of Jacobian matrix.
    $jacobianTranspose = $this->transpose( $jacobianMatrix );

    // Compute J^T * W
    // Where J^T is the tranpose of the Jacobian matrix and W are the weights
    $weightedJacobianIntermediate = $this->multiply( $jacobianTranspose, $weights );

    // Compute (J^T * W) * J.
    // Where J is the Jacobian matrix, and J^T is the tranpose.
    $jacobianIntermediate = $this->multiply( $weightedJacobianIntermediate, $jacobianMatrix );

    // Get the residual error R. This matrix defines how "good" the fit is.
    $errorMatrix = $this->getErrorMatrix( $x, $y, $coefficients );

    // Compute vector $errorIntermediate ( size: numberOfCoefficients x 1) by (J^T*W) * R.
    $errorIntermediate = $this->multiplyWithVector( $weightedJacobianIntermediate, $errorMatrix );
    // Compute the amount of change needed to each coefficient.
    $coefficientsDelta = $this->solve( $jacobianIntermediate, $errorIntermediate );

    // Adjust coefficient by change needed.
    for ( $index = 0; $index < $numberOfCoefficients; ++$index )
        $coefficients[ $index ] += $coefficientsDelta[ $index ];

    // Return new guess at coefficients.
    return $coefficients;
}

/**

```

```

* Compute R-Squared.
*
* Compute Coefficient of determination (R squared) for a set of
* coefficients. This is useful for determining how good a fit the
* coefficients are. (See 5)
*
* @param array $x Array of x data points.
* @param array $y Array of y data points that correspond to x.
* @param array $coefficients Current coefficients to compare against.
* @return float Real value between 0 and 1.
*/
public function getR_Squared( $x, $y, $coefficients )
{
    $errorMatrix = $this->getErrorMatrix( $x, $y, $coefficients );
    $average = array_sum( $y ) / count( $y );
    $totalSumOfSquares = 0;
    $residualSumOfSquares = 0;
    foreach ( $errorMatrix as $index => $error )
    {
        $totalSumOfSquares += pow( $y[ $index ] - $average, 2 );
        $residualSumOfSquares += $error[ 0 ] * $error[ 0 ];
    }
    return 1 - ( $residualSumOfSquares / $totalSumOfSquares );
}
}

```

B. HERDEFINIËRING PAGINAGROEPEN BOL.COM

Bol-label	Paginagroep	Bol-label	Paginagroep
2ndhand	overig	ind	overig
aanbiedingen	overig	indexaspx	overig
account	account	indexhtmbr	overig
account_login	account	indexhtml	overig
account_orders	account	indexhtml20-20	overig
account-aanmaken	account	indexhtml200MGEV-UAK72-21TV1-88APA	overig
acsd	overig	indexhtml20verkoopaccount	overig
actiebal	overig	indexhtmlbr	overig
actiebol	overig	indexhtmlcomo20curar20gripe	overig
adres	overig	indexhtmldix	overig
adres-toevoegen	overig	indexhtmlcom	overig
advent	overig	indexhtmlnl	overig
adventskalender	overig	indexhtmlolcom	overig
afhaalpunten	overig	indexhtmlrtlnl	overig
alle-boeken	overig	indexhtmlSSX20tricky	overig
artikel-niet-meer-leverbaar	overig	indexhtmlyou	overig
bestelling-aanvullen	overig	indkobo20exhtml	overig
bestelling-afgerond	overig	klantenservice	overig
bestelstatus	overig	koken-tafelen	zoeken
betalen	overig	list_2ehands	productcategorieen
betaling-uitvoeren	overig	list_algemeen	productcategorieen
bezorgadres	overig	list_baby	productcategorieen
bezorgadres-toevoegen	overig	list_boeken	productcategorieen
bezorgdetails	overig	list_business-to-business	productcategorieen
bibliotheek	overig	list_cadeaubo	productcategorieen
boeken	zoeken	list_cadeaubon	productcategorieen
brand_2ehands	productcategorieen	list_computer	productcategorieen
brand_algemeen	productcategorieen	list_dier	productcategorieen
brand_baby	productcategorieen	list_duitse-boeken	productcategorieen
brand_boeken	productcategorieen	list_dvd	productcategorieen
brand_computer	productcategorieen	list_elektronica	productcategorieen
brand_dier	productcategorieen	list_engelse-boeken	productcategorieen
brand_dier-tuin-klussen	productcategorieen	list_franse-boeken	productcategorieen
brand_dvd	productcategorieen	list_games	productcategorieen
brand_elektronica	productcategorieen	list_klussen	productcategorieen
brand_games	productcategorieen	list_koken-tafelen	productcategorieen
brand_klussen	productcategorieen	list_mooi-gezond	productcategorieen
brand_koken-tafelen	productcategorieen	list_muziek	productcategorieen
brand_mooi-gezond	productcategorieen	list_nederlandse-boeken	productcategorieen
brand_muziek	productcategorieen	list_service	productcategorieen
brand_service	productcategorieen	list_spaanse-boeken	productcategorieen
brand_speelgoed	productcategorieen	list_speelgoed	productcategorieen
brand_sport-vrije-tijd	productcategorieen	list_sport-vrije-tijd	productcategorieen
brand_tuin	productcategorieen	list_tuin	productcategorieen
brand_wonen	productcategorieen	list_turkse-boeken	productcategorieen
business	overig	list_voordeel	productcategorieen
cadeaubon	overig	list_wonen	productcategorieen
cadeaukaart	overig	menu	overig
cadeauservice	overig	mooi-gezond	zoeken
cartridge	overig	newsletter	overig
catalog	overig	nieuwsbrief	overig
catalogus	overig	nieuwsbrieven	overig
chckout	betalen	nl	overig
checkout	betaald	opruimingindexhtml	overig
checkout_check	betalen	order	overig
checkout_login	login betalen	order_basket	winkelwagen
checkout_payment	overig	Other...	overig
cms_2ehands	productcategorieen	pagina-niet-gevonden	overig
cms_algemeen	productcategorieen	partnerprogramma	overig
cms_baby	productcategorieen	personalization	overig
cms_boeken	productcategorieen	product	producten
cms_bus	productcategorieen	profiel	overig
cms_business-to-business	productcategorieen	recommendations	overig
cms_cadeaubon	productcategorieen	sdd	overig

cms_cadeaubonl	productcategorieen	search_2ehands	zoeken
cms_cadeaukaart	productcategorieen	search_algemeen	zoeken
cms_cart	productcategorieen	search_baby	zoeken
cms_computer	productcategorieen	search_boeken	zoeken
cms_dier	productcategorieen	search_computer	zoeken
cms_dier-tuin-klussen	productcategorieen	search_dier	zoeken
cms_dvd	productcategorieen	search_dvd	zoeken
cms_elektronica	productcategorieen	search_elektronica	zoeken
cms_engelse-boeken	productcategorieen	search_engelse-boeken	zoeken
cms_games	productcategorieen	search_games	zoeken
cms_indexhtml	productcategorieen	search_klussen	zoeken
cms_klantenservice	productcategorieen	search_koken-tafelen	zoeken
cms_klussen	productcategorieen	search_mooi-gezond	zoeken
cms_koken-tafelen	productcategorieen	search_muziek	zoeken
cms_mooi-gezond	productcategorieen	search_nederlandse-boeken	zoeken
cms_muziek	productcategorieen	search_speelgoed	zoeken
cms_nederlandse-boeken	productcategorieen	search_sport-vrije-tijd	zoeken
cms_partnerprogramma	productcategorieen	search_tuin	zoeken
cms_persberichten	productcategorieen	search_wonen	zoeken
cms_servhtml	productcategorieen	service	overig
cms_service	productcategorieen	sinterklaas	overig
cms_speelgoed	productcategorieen	sport-vrije-tijd	overig
cms_sport	productcategorieen	studybooks	overig
cms_sport-vrije-tijd	productcategorieen	trackandtrace	overig
cms_tuin	productcategorieen	undefined	overig
cms_voordeel	productcategorieen	uploads	overig
cms_voorwaarden	productcategorieen	verkopenviabol	overig
cms_wonen	productcategorieen	verkoper	overig
content	overig	verlanglijstje	overig
controle	overig	wachtwoord-instellen	overig
creator_2ehands	productcategorieen	wachtwoord-vergeten	overig
creator_algemeen	productcategorieen	winkelwagentje	winkelwagentje
creator_annuleren-en-retourneren	productcategorieen	wishlist	verlanglijstje
creator_boeken	productcategorieen	wonen	zoeken
creator_duitse-boeken	productcategorieen	adminhtml	overig
creator_dvd	productcategorieen	afhaalpunt-kiezen	overig
creator_engelse-boeken	productcategorieen	apps	overig
creator_muziek	productcategorieen	carnaval	overig
creator_nederlandse-boeken	productcategorieen	creator_baby	productcategorieen
creator_speelgoed	productcategorieen	creator_cd	productcategorieen
doorbots	overig	creator_elektronica	productcategorieen
dsfsdfsdf	overig	creator_undefined	productcategorieen
dvd	overig	creator_wonen	productcategorieen
ebooks	overig	disney-winnie-the-pooh-boeken	overig
electronica	zoeken	errorhtml	overig
elektronica	zoeken	errorjsp	overig
external	overig	indexhtgml	overig
external	overig	indexhtm	overig
facturen	overig	indexhtml3	overig
factuuradres	overig	indexhtmla	overig
factuuradres-toevoegen	overig	indexhtmlbolcombr	overig
fstd_home	overig	indexhtmlr	overig
fstd_jongen	overig	indexhtmspeel20je20wijs	overig
fstd_man	overig	indexingnl	overig
fstd_meisje	overig	internal	overig
fstd_schoen	overig	karcher-ruitenreiniger-window-vac-wv-50-plus	overig
fstd_sint	overig	list_compute	productcategorieen
fstd_vrouw	overig	st_dier-tuin-klussen	productcategorieen
games	overig	list_e	productcategorieen
geen-cookies	overig	list_indexhtml	productcategorieen
harrows	overig	nieuwsbrieve	overig
heli-pack-11l-black	overig	ogone-betaling-fout	overig
home_baby	productcategorieen	pagenotfoundhtml	overig
home_blu-ray	productcategorieen	speelgoed	overig
home_computer	productcategorieen	swww	overig

home_dier	productcategorieen	teunhoevebaars	overig
home_duitse-boeken	productcategorieen	teunhoevenaars	overig
home_dvd	productcategorieen	voordeel	overig
home_electronica	productcategorieen	wandbeugel203720inch	overig
home_engelse-boeken	productcategorieen	wwwpugames	overig
home_franse-boeken	productcategorieen		
home_games	productcategorieen		
home_games-hardware	productcategorieen		
home_gezondheid	productcategorieen		
home_huishouden	productcategorieen		
home_klussen	productcategorieen		
home_koken	productcategorieen		
home_make-up-geuren	productcategorieen		
home_MHP	home_mhp		
home_muziek	productcategorieen		
home_nl-boeken	productcategorieen		
home_spaanse-boeken	productcategorieen		
home_speelgoed	productcategorieen		
home_tafelen	productcategorieen		
home_tuin	productcategorieen		
home_turkse-boeken	productcategorieen		
home_verzorging	productcategorieen		
home_vlaamse-boeken	productcategorieen		