

Tobias Aubele, Mario Fischer

Content-Tracking – aber richtig!

Perfekter Content verdient eine perfekte Messung. Leider unterbleibt die in den meisten Fällen. Es ist auch nicht ganz einfach, wenn man inhaltlich zum Teil sehr unterschiedliche Seiten im Netz stehen hat. Da gibt es auf der einen Seite Produktseiten, die eher wenig Text beinhalten, und auf der anderen Seite z. B. How-to-dos, FAQ, Anleitungen, Erklärseiten oder einen integrierten Blog. Die Webanalyse-Tools wie Google Analytics werten für alle diese URLs die Besuchsdauer aus. Am Ende hat man eine Liste mit mehreren Tausend URLs und mehreren Tausend Zeitangaben dahinter. Was soll man damit anfangen? Genau – erst mal gar nichts. Man nimmt es zur Kenntnis. Recherchiert wird meist nur in Einzelfällen, wenn die Zeiten besonders stark von der Norm abweichen. Aber wie kann man beurteilen, ob und wie intensiv der lesenswerte Textabschnitt, der sog. „Primary Content“, überhaupt von Besuchern konsumiert wird? Mit ein wenig Engagement und einigen Tricks geht das sehr wohl – zumindest näherungsweise. Wie, das zeigen Tobias Aubele und Mario Fischer in diesem Beitrag. Nachahmung wird empfohlen.

Google Analytics ist im Standard ungenau!

Google Analytics liefert in Bezug auf Contentbewertung eine Vielzahl an Kennzahlen. Leider unterliegen all diese Kennzahlen einem technisch bedingten Messfehler, sofern Analytics ohne Modifikationen in die Site eingebunden wurde. Startet z. B. ein Besucher um 10:00 Uhr eine Sitzung und schaut sich drei Seiten an, so wird die Sitzungsdauer durch die Subtraktion des letzten Hits (Seitenaufruf um 10:10 Uhr; siehe Abbildung 1) vom ersten Hit (hier der Seitenaufruf) ermittelt. Obwohl der Besucher möglicherweise auf der letzten Seite weitere fünf Minuten verweilt und bspw. den Content komplett liest, wird diese Zeit bzw. die Bedeutsamkeit dieses Contents nicht gewertet und tatsächlich sogar auf 0 gesetzt. Die Seite 3 wird damit qualitativ extrem unterbewertet.

Besonders dramatisch: Sofern der Besuch nur aus einem einzigen Seitenaufruf besteht und der Nutzer keine weiteren Seiten durch das Anklicken eines Links lädt, bekommen Webserver und das Tracking-Tool keine weitere Seitenanforderung. Es gibt also kein abschließendes Signal. Und auch hier wird die Besuchsdauer des Nutzers auf 0 Sekunden gesetzt. Der (Ein-Seiten-)Besuch wird als Bounce (Absprungrate = 100 %) gewertet, obwohl möglicherweise der

komplette Content gelesen und die Nutzungsintention perfekt befriedigt wurde! Dadurch werden völlig falsche Signale an die Verantwortlichen oder Redakteure gesendet: „Der Nutzer findet den Content wohl nicht interessant!“ Gerade bei Blogs mit Newsletter ist dies von großem Nachteil, da oft nur der im Newsletter beworbene Beitrag gelesen wird. Es sieht aber messtechnisch so aus, als wären dies die Verlierer-Seiten.

Eine Lösung ist die Integration von Ereignissen auf der Seite, um das tatsächliche Nutzerverhalten besser zu messen. Ein Ereignis kann sowohl seitens des Nutzers erfolgen, wie z. B. das Starten eines Videos oder das Klicken auf einen Button zum Download eines PDFs, als auch zeitgesteuert ausgelöst werden. Hier fragt ein Skript in Zeitintervallen an, ob die Seite noch im Browser aktiv ist. Sobald ein Ereignis und damit eine Interaktion (Engagement) auf der Seite stattfindet, hat das Tracking-System einen weiteren Zeitstempel und kann dadurch die Zeit „korrekter“ ermitteln. Findet auf der dritten Seite also um 10:15 Uhr ein Ereignis statt (Nutzer lädt z. B. ein PDF), so wird dieser Zeitstempel als letzter Endpunkt gesetzt und die gleiche Sitzung dauert nun 15 Minuten (Abbildung 2).

Wie angesprochen, ermittelt das System die Zeit als Differenz zwischen einzelnen Hits

DER AUTOR



Tobias Aubele ist Professor für E-Commerce, insbesondere Conversion-Optimierung und Usability an der Hochschule Würzburg-Schweinfurt. Darüber hinaus berät er KMU im Bereich Webanalytics & Website-Optimierung.

DER AUTOR



Mario Fischer ist Professor für E-Commerce an der Hochschule Würzburg-Schweinfurt und Herausgeber der Website Boosting. Er beschäftigt sich u. a. seit 1996 intensiv mit dem Thema Suchmaschinen.

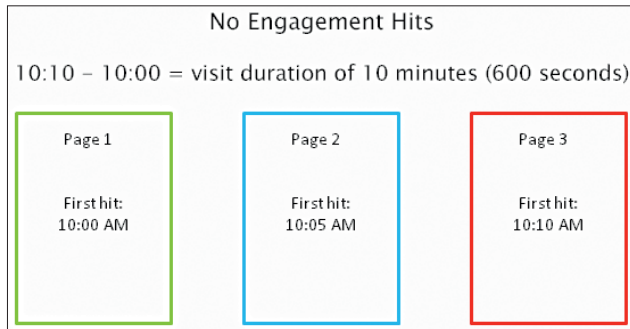


Abb. 1: Ermittlung der Besuchszeit bei drei Seitenaufrufen
(Quelle: Google; <http://einfach.st/ga43>)

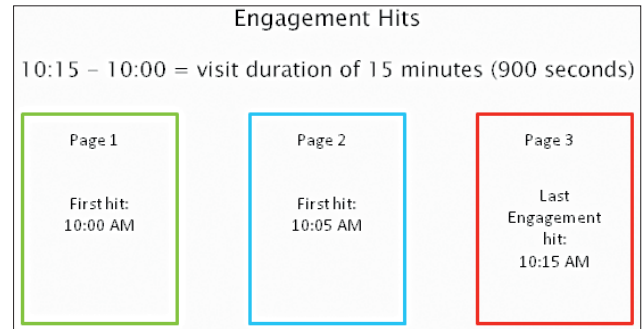


Abb. 2: Ermittlung der Besuchszeit unter Berücksichtigung von Ereignissen
(Quelle: Google; <http://einfach.st/ga43>)

– meist Seitenaufrufen – und schließt nach einer Inaktivität von 30 Minuten automatisch die Session, da der Browser nicht an Google Analytics zurückmeldet, wenn der Nutzer die Seite schließt. Bei Seiten mit einem hohen Anteil an Videocontent empfiehlt es sich daher, die standardmäßige Sitzungseinstellung von 30 Minuten zu erhöhen (Abbildung 3).

Der Durchschnitt lügt! Was tut der Nutzer wirklich?

Um die Nützlichkeit des Contents beurteilen zu können, sind zwei Aspekte von großer Bedeutung:

1. Ermittlung von Nutzungsinformationen, d. h., wie stark wurde der Content tatsächlich konsumiert. Fragen hierbei sind: Wie lange war der Nutzer auf der Seite? Wie tief wurde gescrollt? Waren bestimmte Content-

bereiche bzw. Elemente überhaupt im sichtbaren Bereich?

2. Verteilung der Daten, d. h., welche Bandbreite haben die erhobenen Daten. Eine durchschnittliche Besuchszeit von einer Minute wird auch erreicht, wenn neun Besucher nur eine Sekunde auf der Seite waren und ein Besucher 9:51 Minuten mit dem Content interagierte. Der Durchschnitt lügt – immer!

Ultraschnelles
High-Performance
SSD-Webhosting mit nginx

„Ein Jäger schießt links am Hasen vorbei, der andere rechts. Der Durchschnitt zeigt aber genau zwischen die Augen und signalisiert Erfolg.“

So ist der Website-Content-Bericht (Menüpunkt „Verhalten“) in Google Analytics mit seinen Metriken (Abbildung 4) nur ein Indikator für die Leistungsfähigkeit der einzelnen Seite. Sofern die analysierte Seite die letzte Seite des Besuches war, ist diese Kennzahl sogar komplett wertlos.

Definition von Ereignissen in Google Analytics

Damit in Google Analytics ein Ereignis getrackt wird, ist eine Anpassung im Tracking-Code notwendig. Die einfachste und transparenteste Art des Trackings wird durch den Einsatz des Google-Tag-Managers ermöglicht (<https://tagmanager.google.com>). Neben der hohen Flexibilität ist insbesondere die Versionierung und damit zeitbezogene Dokumentation des Tracking-Schemas ein enormer Vorteil: Die meisten Anwender wissen nicht mehr, mit welcher Variante des Tracking-Codes die Daten im Vorjahr erhoben wurden und verzerren damit ggf. unbewusst die Analysen.

Um ein Ereignis an Google Analytics zu übermitteln, benötigt man ein sog. Ereignis-Tag nebst zugehörigem Trigger, d. h. eine Definition, wann und wo das Tag ausgelöst werden soll. Soll z. B. nach 30 Sekunden abgefragt werden, ob die Seite noch im Browser aktiv ist (und demnach die Site vom Besucher nicht bereits verlassen wurde), so bedarf es eines zeitgesteuerten Triggers. Nach weiteren 30 Sekunden bzw. frei wählbaren Intervallen kann

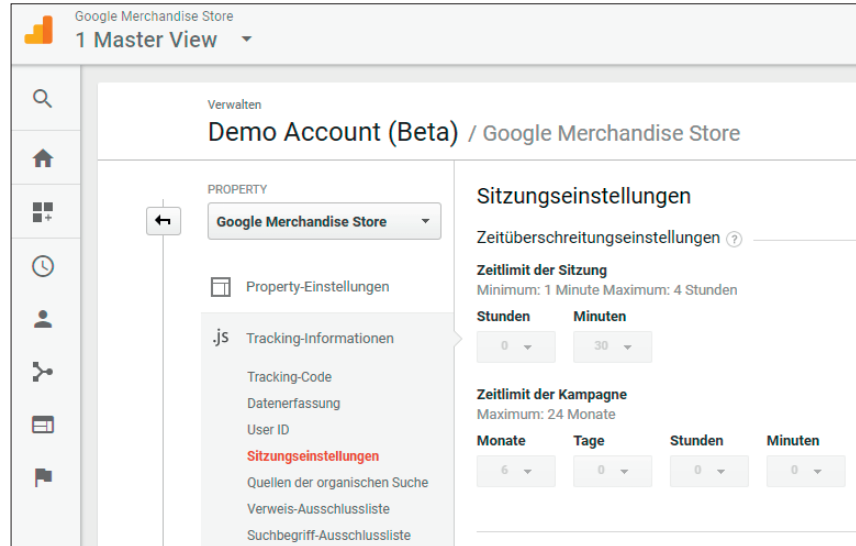


Abb. 3: Sitzungseinstellungen in der Property-Verwaltung

Seite	Seitenaufufe	Einzelne Seitenaufufe	Durchschn. Zeit auf der Seite	Einstiege	Absprungrate	% Ausstiege	Seitenwert
	4.454 <small>% des Gesamtwerts: 100,00 % (4.454)</small>	3.293 <small>% des Gesamtwerts: 100,00 % (3.293)</small>	00:00:46 <small>Durchschn. für Datensicht: 00:00:46 (0,00 %)</small>	1.176 <small>Durchschn. für Gesamtwert: 100,00 % (1.176)</small>	52,38 % <small>Durchschn. für Datensicht: 52,38 % (0,00 %)</small>	26,40 % <small>Durchschn. für Datensicht: 26,40 % (0,00 %)</small>	7,04 \$ <small>% des Gesamtwerts: 100,00 % (7,04 \$)</small>
1. /home	976 (21,91 %)	725 (22,02 %)	00:00:48	563 (47,87 %)	48,49 %	39,04 %	1,12 \$ (15,93 %)
2. /google+redesign/shop+by+brand/youtube	415 (9,32 %)	349 (10,60 %)	00:00:47	330 (28,06 %)	61,56 %	56,39 %	0,16 \$ (2,28 %)
3. /basket.html	245 (5,50 %)	101 (3,07 %)	00:00:36	7 (0,60 %)	42,86 %	14,29 %	21,54 \$(305,90 %)
4. /signin.html	141 (3,17 %)	86 (2,61 %)	00:00:29	16 (1,36 %)	18,75 %	11,35 %	21,83 \$(310,05 %)

Abb. 4: Website-Content mit Informationen zu Seitenaufrufen und durchschnittlicher Besuchszeit

dann die nächste Abfrage erfolgen, ob der Nutzer immer noch da ist (Abbildung 5). Im ersten Moment klingt dies vielleicht seltsam, dass permanent nachgefragt werden soll, aber Google Analytics bekommt wie oben erwähnt diese Information sonst nicht.

Nachdem ein Trigger definiert wurde, kann ein Universal-Analytics-Tag diesen Trigger aufnehmen und ein Ereignis an Analytics übergeben. Neben dem Tracking-Typ „Ereignis“ sowie der Tracking-ID kann die Kategorie sowie Aktion fest vorgegeben werden und ist in der späteren Analyse dann auch die Bezeichnung in Analytics. Die Variable „URL“ im Label ermöglicht, dass das Ereignis eindeutig den Seiten zuordenbar ist. Weiterhin ist aus Datenschutzgründen die IP-Anonymisierung zu aktivieren. Als Trigger wird der soeben definierte Trigger „Timer 30s“ hinterlegt (Abbildung 6). Diese Kombination aus Trigger und

Ereignis-Tag kann nun für weitere Kombinationen angelegt werden (z. B. nach 60 Sekunden, 120 Sekunden usw.).

Tipp: Die Tracking-ID könnte einmalig als Variable definiert werden und sich die weiteren Definitionen des Trackings auf diese Variable beziehen. Damit ist das Tracking-Konzept für mehrere Domains einsetz- bzw. kopierbar: Einfach die Tracking-ID bei der neuen Property austauschen und schon ist das Konzept mehrfach einsetzbar.

In Google Analytics ist demnach die Zeit durch die Tag-Definition direkt ablesbar. Nach 30 Sekunden waren z. B. von den zehn Sitzungen noch fünf aktiv, nach 60 Sekunden nur noch vier. Bis zum Trigger 480 Sekunden haben bereits acht Besucher die Seite verlassen (Abbildung 7).

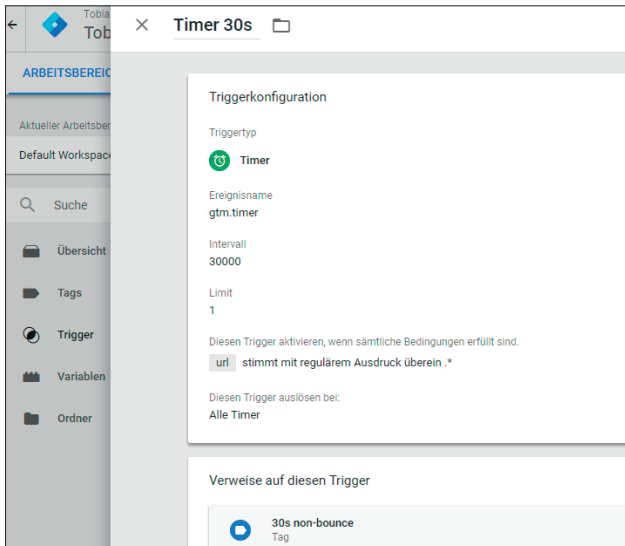


Abb. 5: Zeitgesteuerter Timer auf allen Seiten einer Site

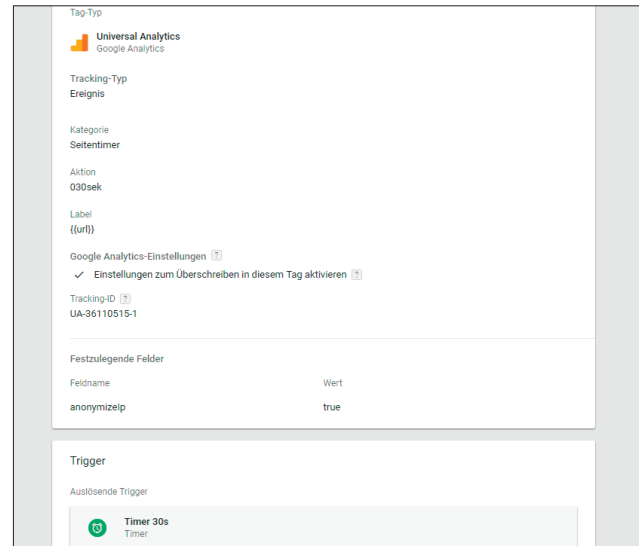


Abb. 6: Tag-Definition für ein Ereignis

Ein wichtiger Hinweis: Je Session stehen nur 500 Hits zur Verfügung bei 10 Mio. Hits pro Property je Monat, d. h., diese Grenze im Tracking-Konzept sollte man unbedingt berücksichtigen. Weitere Hits je Session werden darüber hinaus

nicht mehr erfasst. Weiterhin stehen die 500 Hits erst im Zeitverlauf zur Verfügung: Initial werden 20 Hits zur Verfügung gestellt und je Sekunde um 2 Hits erweitert, d. h., bei der Ereignisdefinition sollte man nicht zu übermütig werden.

Tipp: Im Web sind diverse Schritt-für-Schritt-Anleitungen von (zeitgesteuerten) Ereignissen abrufbar, z.B. hier das Blog von Rocket-Internet (<http://einfach.st/rocket>).

Ultraschnelles
High-Performance
SSD-Webhosting mit nginx



Adjusted Bounce-Rate – Non-Interaktion-Kennzeichnung

Die kontinuierliche Abfrage von zeitgesteuerten Ereignissen liefert nicht nur eine genauere Time on Site, sondern ermöglicht, die Verteilung der Besuchszeiten zu analysieren. Ausreißer in den statistischen Daten können dadurch erkannt werden. Weitere Kennzahlen wie Bounce-Rate sind damit ebenfalls betroffen. Ein unscheinbares Feld in der Tag-Konfiguration hat hierbei einen großen Einfluss – Treffer ohne Interaktion (Abbildung 8). Wird der Wert auf „Wahr“ gesetzt, wird eine Sitzung mit nur einer angesehenen Seite, die mit Ereignissen ohne Interaktionen versehen ist, als Absprung gezählt – selbst wenn der Nutzer während der Sitzung ein Ereignis auslöst. Umgekehrt bewirkt das Auslassen dieser Option, dass eine Sitzung, bei der nur eine einzige Seite mit Ereignis-Tracking besucht wird, nicht als Absprung gezählt wird, wenn der Nutzer während derselben Sitzung ein Ereignis auslöst (<http://einfach.st/ga44>).

Das heißt, sofern die Webseite nach 30 Sekunden noch im Browser aktiv ist, wird die Bounce-Rate auf 0 % und die Time on Site auf 30 Sekunden gesetzt. Diese Sitzung wird damit nicht zum Absprung, obwohl der Besucher ggf. nur diese eine Seite während seines Besuches aufruft („adjusted Bounce Rate“).

Contentbewertung auf URL-Ebene

Der Export von Ereignissen aus Google Analytics in Excel ermöglicht eine detaillierte Auswertung der Verweildauer auf URL-Ebene. Über die Seo-Tools für Excel (umfassender Beitrag dazu siehe in der letzten Ausgabe #44 der Website Boosting) können die Ereignisse direkt adressiert werden (Abbildung 9). Bei den Metriken ist dies das Feld ga:totalEvents (direkte Eingabe in das Feld Metrics bzw. Auswahl

Ereignisaktion	Ereigniskategorie	Ereignisse gesamt	Eindeutige Ereignisse
		20 % des Gesamtwerts: 13,61 % (147)	20 % des Gesamtwerts: 40,00 % (50)
1. 30sek	Seitentimer	5 (25,00 %)	5 (25,00 %)
2. 60sek	Seitentimer	4 (20,00 %)	4 (20,00 %)
3. 120sek	Seitentimer	3 (15,00 %)	3 (15,00 %)
4. 240sek	Seitentimer	3 (15,00 %)	3 (15,00 %)
5. 360sek	Seitentimer	3 (15,00 %)	3 (15,00 %)

Abb. 7: Ereignisse und deren Metriken



Abb. 8: Treffer ohne Interaktion

über den Link „Select“) unter Berücksichtigung der Dimensionen ga:eventCategory, ga:eventLabel sowie ga:eventAction. Mittels eines Filters auf den „Seitentimer“, der gewählten Bezeichnung im Tag, kann die Auswahl gleich direkt im Import reduziert werden.

Mittels einer zweiten API-Abfrage (Metrik „ga:uniquePageviews“ mit der

Dimension „ga:pagePath“) können für jede Seite die entsprechenden einzigartigen Seitenaufrufe ermittelt werden (Abbildung 10). Über die INDEX-Funktion in Excel können die Ergebnisse aus der Event-Abfrage und den Pageviews der zweiten Abfrage kombiniert werden.

Die Formel in Zelle E20 aus Abbildung 10 lautet:
 =WENN(ISTNV(INDEX(\$D\$2:\$D\$15; VERGLEICH(\$D20&" "&\$E\$19;\$E\$2:\$E\$15;0)));0;INDEX(\$D\$2:\$D\$15;VERGLEICH(\$D20&" "&\$E\$19;\$E\$2:\$E\$15;0)))

	A	B
1	ga:eventCategory	ga:eventLabel
2	Seitentimer	http://www.tobias-aubele.de/
3	Seitentimer	http://www.tobias-aubele.de/
4	Seitentimer	http://www.tobias-aubele.de/
5	Seitentimer	http://www.tobias-aubele.de/
6	Seitentimer	http://www.tobias-aubele.de/
7	Seitentimer	http://www.tobias-aubele.de/
8	Seitentimer	http://www.tobias-aubele.de/in
9	Seitentimer	http://www.tobias-aubele.de/le
10	Seitentimer	http://www.tobias-aubele.de/le
11	Seitentimer	http://www.tobias-aubele.de/pr
12	Seitentimer	http://www.tobias-aubele.de/pr
13	Seitentimer	http://www.tobias-aubele.de/pr
14	Seitentimer	http://www.tobias-aubele.de/pr
15	Seitentimer	http://www.tobias-aubele.de/pr
16		
17		
18		
19	ga:hostname	ga:pagePath
20	www.tobias-aubele.de	/

Abb. 9: Definitionen der Analytics-API-Anfrage mittels der Seo-Tools

=WENN(ISTNV(INDEX(\$D\$2:\$D\$15;VERGLEICH(\$D20&"_&E\$19;SE\$2:SE\$15;0)));0;INDEX(\$D\$2:\$D\$15;VERGLEICH(\$D20&"_&E\$19;SE\$2:SE\$15;0)))

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ga:eventCategory	ga:eventLabel	ga:event	ga:totalEvents	KEY											
2	Seitentimer	http://www.tobias-aubele.de/	030sek	501	http://www.tobias-aubele.de/_030sek											
3	Seitentimer	http://www.tobias-aubele.de/	060sek	260	http://www.tobias-aubele.de/_060sek											
4	Seitentimer	http://www.tobias-aubele.de/	120sek	121	http://www.tobias-aubele.de/_120sek											
5	Seitentimer	http://www.tobias-aubele.de/	240sek	50	http://www.tobias-aubele.de/_240sek											
6	Seitentimer	http://www.tobias-aubele.de/	360sek	12	http://www.tobias-aubele.de/_360sek											
7	Seitentimer	http://www.tobias-aubele.de/	480sek	2	http://www.tobias-aubele.de/_480sek											
8	Seitentimer	http://www.tobias-aubele.de/impressum-date	030sek	50	http://www.tobias-aubele.de/impressum-datenschutz/_030sek											
9	Seitentimer	http://www.tobias-aubele.de/lehre/	030sek	14	http://www.tobias-aubele.de/lehre/_030sek											
10	Seitentimer	http://www.tobias-aubele.de/lehre/	060sek	12	http://www.tobias-aubele.de/lehre/_060sek											
11	Seitentimer	http://www.tobias-aubele.de/publikationen/	030sek	200	http://www.tobias-aubele.de/publikationen/_030sek											
12	Seitentimer	http://www.tobias-aubele.de/publikationen/	060sek	150	http://www.tobias-aubele.de/publikationen/_060sek											
13	Seitentimer	http://www.tobias-aubele.de/publikationen/	120sek	121	http://www.tobias-aubele.de/publikationen/_120sek											
14	Seitentimer	http://www.tobias-aubele.de/publikationen/	240sek	5	http://www.tobias-aubele.de/publikationen/_240sek											
15	Seitentimer	http://www.tobias-aubele.de/publikationen/	360sek	1	http://www.tobias-aubele.de/publikationen/_360sek											
16																
17																
18																
19	ga:hostname	ga:pagePath	ga:uniquePfad		030sek	060sek	120sek	240sek	360sek	480sek	kum. Absprung					
20	www.tobias-aubele.de	/	620	http://www.to	501	260	121	50	12	2	19%	58%	80%	92%	98%	100%
21	www.tobias-aubele.de	/impressum-datenschutz/	120	http://www.to	50	0	0	0	0	0	58%	100%	100%	100%	100%	100%
22	www.tobias-aubele.de	/lehre/	17	http://www.to	14	12	0	0	0	0	18%	29%	100%	100%	100%	100%
23	www.tobias-aubele.de	/publikationen/	214	http://www.to	200	150	121	5	1	0	7%	30%	43%	98%	100%	100%
24	www.tobias-aubele.de	/kontakt/	7	http://www.to	0	0	0	0	0	0	100%	100%	100%	100%	100%	100%

Abb. 10: Analyse der Verweildauer auf URL-Ebene

Daraus kann der kumulierte Absprung auf URL-Ebene ermittelt werden (Formel: $=($C20-E20)/$C20$). Für die Homepage ergibt sich folgendes Bild:

- » Von 620 Seitenaufrufen sind nach 30 Sekunden noch 501 Browseraktionen aktiv.
- » Nach 360 Sekunden sind es nur noch 12.

Das heißt, nach 60 Sekunden sind bereits 58 % der Besucher nicht mehr auf der Seite aktiv. Theoretisch können weitere statistische Funktionen wie Quartil, Median und klassierte Mittel-

Ultraschnelles
High-Performance
SSD-Webhosting mit **nginx**

werte angewendet werden, welche eine weitere Beurteilung der Seite ermöglichen. Weiterhin ermöglicht die bedingte Formatierung von Zellen in Excel eine visuelle Einschätzung der Seitenperformance (grün = zufriedenstellend, rot = verbesserungswürdig).

Contentengagement – ist der Tab im sichtbaren Bereich?

Eine perfekte Ergänzung zur Time on Page mittels Events ist die Ableitung des Nutzerengagements in Bezug auf den aktiven Tab im Browser. Das heißt, sobald der Nutzer die Seite nur in einem nicht-aktiven Tab geladen hat („Link in neuem Tab öffnen“), soll die Zeit nicht gezählt werden. Im Browser muss demnach die Seite im aktiven Fenster sein und Nutzerinteraktionen wie Mausbewegung oder Tastatureingaben müssen vollzogen werden. Simo Ahava hat die Ermittlung des Nutzerengagements in seinem Blog mittels zweier Methoden detailliert erläutert (<http://einfach.st/simo3>).

Über einen berechneten bzw. über einen benutzerdefinierten Messwert kann somit ein direkter Vergleich zwischen aktiver und durchschnittlicher Zeit auf der Seite gezogen werden. Dabei können erhebliche Diskrepanzen zutage treten. So freut man sich ggf. über eine Verweildauer auf der Galerieseite von 46 Sekunden, was jedoch in Wahrheit nur sieben Sekunden echte Aufmerksamkeit sind, die restliche Zeit ist die Seite gar nicht im Fokus.

Tipp: Zur optimalen Analyse setzt man die neue Metrik „Average Engagement“ in einem benutzerdefinierten Bericht in Relation zu bestehenden Kennzahlen (Abbildung 11). In diesem Berichtstyp besteht komplette Flexibilität bei gleichzeitiger Strukturierung in diversen Tabs und Darstellungsalternativen. Diese Berichte sind generell

Seite ?	Seitenaufufe ?	Absprungrate ?	Average Engagement ↓	Durchschn. Zeit auf der Seite ?
	147 % des Gesamtwerts: 100,00 % (147)	15,38 % Durchn. für Datensicht: 15,38 % (0,00 %)	00:00:04 % des Gesamtwerts: 100,00 % (00:00:04)	00:02:37 Durchn. für Datensicht: 00:02:37 (0,00 %)
1. /	62 (42,18 %)	18,18 %	00:00:08(171,53 %)	00:02:42
2. /gallery/	3 (2,04 %)	0,00 %	00:00:07(168,14 %)	00:00:46

Abb. 11: Messung von Nutzerengagement

<input type="checkbox"/>	Ereignisaktion ?	Ereignisse gesamt ↓	Eindeutige Ereignisse ?
		59.886 % des Gesamtwerts: 79,17 % (75.645)	46.930 % des Gesamtwerts: 78,74 % (59.599)
<input type="checkbox"/>	1. 25%	26.255 (43,84 %)	19.409 (41,36 %)
<input type="checkbox"/>	2. 50%	18.115 (30,25 %)	14.368 (30,62 %)
<input type="checkbox"/>	3. 75%	12.602 (21,04 %)	10.474 (22,32 %)
<input type="checkbox"/>	4. 100%	2.914 (4,87 %)	2.679 (5,71 %)

Abb. 12: Scrollverhalten der Besucher

sehr hilfreich, da Metriken und Dimensionen beliebig miteinander kombinierbar sind.

Scroll-Tracking – schaut sich der Nutzer auch das Wichtige an?

Weitere Erkenntnisse über die Qualität des Contents lassen sich durch das Tracking des Scrollverhaltens ableiten; diese Erweiterung des Tracking-Konzepts ist absolut empfehlenswert. Sofern konversionsfördernde Elemente wie z. B. Erklärvideos und Testimonials unter der Falz eingebettet sind, ist es wichtig zu wissen, ob diese überhaupt seitens der Besucher mittels Scrollen in den Fokus kommen. Scroll-Tracking kann über mehrere Möglichkeiten in Google Analytics integriert werden; das wohl bekannteste Plug-in ist von Rob Flaherty (<http://scrolldepth.parsnip.io/>). Mittels eines benutzerdefinierten HTML-Tags kann das Skript über den Tag-Manager bzw. über direkt in WordPress mittels des Plug-ins „WP Scroll Depth“ integriert werden. Sofern die Website nicht über jQuery größer Version 1.7 verfügt, existieren im Netz weitere Skriptalternativen. In Abhängigkeit vom Scrollverhalten werden auf URL-Ebene Prozentwerte ausgegeben.

So hatten von 32.475 Seitenaufrufe 26.255 Aufrufe über 25 % der gesamten Seitenhöhe im sichtbaren Bereich. Letztlich wurde aber nur bei 2.914 Seiten bis zum Ende gescrollt (Abbildung 12).

Weiterhin kann mittels Scroll-Tracking geprüft werden, ob bestimmte Contentelemente in den sichtbaren Bereich gelangten (z. B. Kundenrezensionen) und wie viele Pixel beim Scrollen jeweils bewegt wurden. Eine Analyse auf URL-Ebene in Kombination mit der oben erläuterten Verweildauer ermöglicht, schlecht performende Seiten einem qualitativen Audit zu unterziehen und ggf. Contenttests durchzuführen (bspw. mit Google Optimize in einer B-Variante Textelemente hinsichtlich ihrer Position und Größe zu verschieben). In Summe lassen sich damit viele Fragen zum Content beantworten:

- » Mit welchen Suchbegriffen kommen die Nutzer auf die Seite (bspw. über die Google Search Console)?
- » Wie weit scrollen sie – welcher Content ist damit nicht im sichtbaren Bereich und sollte weiter oben positioniert werden?
- » Wie lange bleiben sie auf der Seite und können Sie damit die zur Verfügung gestellten Informationen in der Zeit auch gelesen haben?

Tipp: Unter <http://einfach.st/lunametrics> werden komplette „Rezepte“ (spezifische Tracking-Pakete) für den Google-Tag-Manager zur Verfügung gestellt, welche direkt über einen Container importiert werden können. Dadurch werden passende Tags, Trigger sowie Variablen für eine Tracking-Anwendung (Cookie-Management, Scroll-Tracking, Events, um ausgehende Links zu tracken, etc.) gesammelt zur Verfügung gestellt.

Wie ermittelt man die Anzahl Wörter einer Seite?

Wenn nun klar ist, wie lange jemand tatsächlich auf einer Seite verweilt, muss man diesen Werten den Umfang des konsumierbaren Textes gegenüberstellen. Auf den ersten Blick geht das recht einfach mit diversen SEO-Tools wie dem Screaming Frog (Abbildung 13) oder den SEO-Tools für Excel (Abbildung 14), die einfach die Anzahl der Wörter pro URL ermitteln. Aufmerksame Leser werden bemerken, dass der Frog jeweils drei bis vier Wörter weniger ermittelt, aber das ist für die vorliegenden Zwecke unwichtig.

Aber was haben wir davon? Richtig, nichts. Denn jede normale Webseite hat ein mehr oder weniger großes Set an Wörtern, die den eigentlichen Text umgeben – die sog. Boilerplate (Abbildung 15). Je mehr solcher Textbausteine oder Wörter im Head, Footer, einer Sidebar oder an anderen Stellen jeweils auf allen Seiten rumschwirren, desto ungenauer wird die einfache Ermittlung mittels der Wort-Zählfunktion von Tools.

Wie ermittelt man den echten „Primary Content“?

Die schlechte Nachricht ist, dass man dies manuell tun muss und automatisierte Tools hier ins Leere greifen,

Address	Content	Sta	Size	Word Count	Text Ratio
1 https://www	text/html; charset=UTF-8	200	703	250	4,85
2 https://www	text/html; charset=UTF-8	200	214	411	7,24
3 https://www	text/html; charset=UTF-8	200	18	255	4,81
4 https://www	text/html; charset=UTF-8	200	143	334	5,84
5 https://www	text/html; charset=UTF-8	200	386	947	12,19
6 https://www	text/html; charset=UTF-8	200	532	397	7,07
7 https://www	text/html; charset=UTF-8	200	164	395	7,04
8 https://www	text/html; charset=UTF-8	200	387	331	6,03
9 https://www	text/html; charset=UTF-8	200	998	253	4,68

Abb. 13: Die Ermittlung der Wortanzahl für URL im Screaming Frog

B	C
v.w. ...	253
v.w. ...	414
v.w. ...	258
v.w. ...	337
v.w. ...	952
v.w. ...	405
v.w. ...	404
v.w. ...	334
v.w. ...	256
v.w. ...	55

Abb. 14: Die Ermittlung der Wortanzahl für URL in den SEO-Tools für Excel

Boilerplate!

Primary Content!

Abb. 15: Bei der Ermittlung des richtigen Textumfangs muss zuerst die Boilerplate entfernt werden

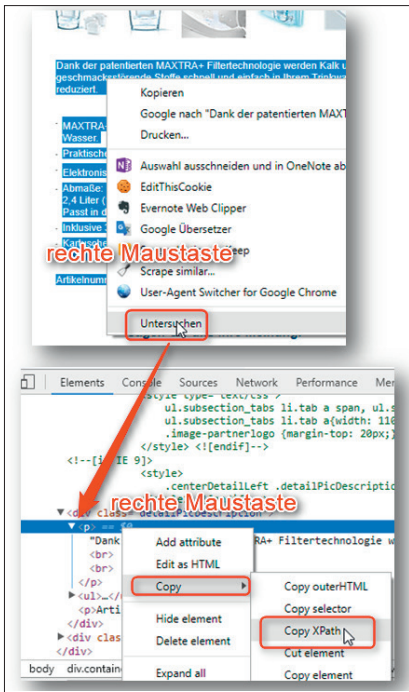


Abb. 16: Text markieren, rechte Maustaste, „Untersuchen“, rechte Maustaste auf den unten blau markierten Bereich, „Copy“, „Copy XPath“ – fertig

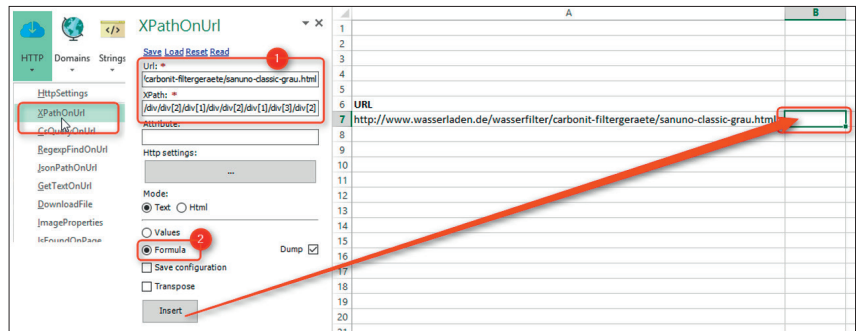


Abb. 17: Der Formeleditor der SEO-Tools baut alles bequem zusammen

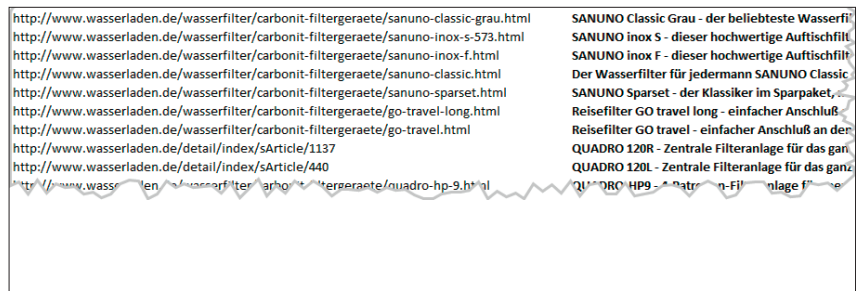


Abb. 18: Ist die Formel erst einmal generiert, kann man die Texte beliebiger URLs holen

weil jede Website strukturell bzw. von der Programmierung her anders aufgebaut ist. Die gute Nachricht ist, dass es in der Regel sehr einfach zu bewerkstelligen ist. Rufen Sie einfach eine typische Produktseite, einen Blogbeitrag oder eine andere Webseite auf und markieren den eigentlich interessanten, unikalsten Text mit der Maus. Klicken Sie nun mit der rechten Maustaste und wählen im Chrome-Browser „Untersuchen“ (Abbildung 16).

Hinweis: Für den Firefox-Browser geht das mit der Erweiterung Firebug oder einem speziellen Plug-in dafür unter <http://einfach.st/xpfirefox>. Eine gute Anleitung für den IE und den Safari findet man unter <http://einfach.st/xpathbrowser>.

Es öffnet sich unten das Entwicklerfenster und der eben markierte Textabschnitt wird als Zeile im sog. DOM blau markiert. Man kann sich den DOM (Document Object Model) einfach als einen Strukturbaum vorstellen, der alle einzelnen HTML-Elemente enthält. Auf diese kann man dann „von außen“ über

die Adresse, den „XPath“, zugreifen. Um diesen in die Zwischenablage zu kopieren, fährt man mit dem Mauszeiger einfach die blau markierte Zeile an und wählt mit einem Rechtsklick „Copy“ und dann „Copy XPath“ aus. In der Zwischenablage liegt nun ein String, der vom Muster her in etwa so aussieht bzw. je nach Webseite im Detail natürlich anders:

```
/html/body/div[1]/div[5]/div[3]/div[1]/div[4]/p[1]
```

Damit ist der erste Schritt auch schon erledigt.

Nun bringt man diesen Pfad noch mit den entsprechenden URLs zusammen. Mit den SEO-Tools für Excel geht das ebenfalls recht einfach.

Man wählt unter dem Menü „HTTP“ die Funktion „XPathOnUrl“ aus und es öffnet sich ein Konfigurationsfenster, wie in Abbildung 17 gezeigt wird. Dort trägt man zunächst die entsprechende URL und den XPath aus der Zwischenablage ein (Markierung 1). Wichtig ist, von der Grundeinstellung „Values“ zu „Formula“ zu wechseln. Dann aktiviert man die Zelle, in die die Formel eingefügt werden soll – im Beispiel ist das

B7. Ein Klick auf „Insert“ fügt die Formel ein und holt auch gleich den vorhin selektierten Text in die Zelle.

Klickt man auf die Zelle (B7), sieht man die eigentliche, eben per Editor zusammengesetzte Formel, die im Beispiel so aussieht:

```
=Dump(XPathOnUrl(„http://www.wasserladen.de/wasserfilter/carbonit-filtergeraete/sanuno-classic-grau.html“;“/html/body/div[1]/section/div/div/div[2]/div[1]/div/div[2]/div[1]/div[3]/div[2]“;;“text“))
```

Darin ist nun sowohl die URL (roter Text) als auch der XPath (blauer Text) sowie einige Parameter dahinter, die hier keine Rolle spielen. In der Zelle A7 steht im Beispiel noch die ursprüngliche URL. Jetzt markiert man einfach in der Formel in B7 die gesamte URL einschließlich der Anführungszeichen und ersetzt den Eintrag mit dem Zellbezug auf A7:

```
=Dump(XPathOnUrl(A7;“/html/body/div[1]/section/div/div/div/div[2]/div[1]/div/div[2]/div[1]/div[3]/div[2]“;;“text“))
```

Das Ergebnis bleibt gleich, aber die Formel ist jetzt dynamisch an die

Zelle links daneben gebunden (A7). Wenn man nun mehrere URLs in Spalte A untereinander stellt, kann man die Formel in B7 einfach nach unten ziehen. Jetzt holt Excel nach und nach alle Textinhalte der URL, die im gleichen XPath stehen (Abbildung 18).

Wenn man unterschiedliche Vorlagen auf der Website hat, z. B. einen zusätzlichen Blog, wiederholt man diesen Vorgang einfach für die jeweils andere DOM-Struktur. Also den eigentlichen Text dort markieren, den XPath kopieren und dann über das Set weitere strukturgleiche URLs holen.

Da man nun alle Texte aus den Primary-Content-Bereichen in Zellen gezogen hat, ist die weitere Auswertung einfach. Über die Funktion =CountWords(A7), die mit den SEO-Tools für Excel ebenfalls integriert wird (über das Menü unter „Strings“), steht eine einfache Wortzählung einer Zelle zur Verfügung. Man adressiert also einfach die Zelle, in der man per XPath den Text geholt hat, und erhält die Anzahl Wörter.

Hinweis: Die Funktion „=CountWords“ darf nicht mit der Funktion „=WordCount“ verwechselt werden. Letztere zählt wie weiter oben beschrieben die Anzahl Wörter auf einer Webseite. Die erste Formel zählt die Wörter in einer Zelle!

Wenn man nun weiß, dass normale Menschen etwa mit einer Geschwindigkeit von 220 Wörtern pro Minute lesen, kann man die eben ermittelte Anzahl der Wörter aus den Primary-Content-Bereichen, z. B. 1.620 Wörter, durch Division dazu in Beziehung setzen. Wer diese Seite komplett durchlesen möchte, braucht also in etwa grob 7,4 Minuten (1.600 Wörter/220 Wörter).

Die oben ermittelte deutlich exaktere Zeitmessung pro URL via Ana-

lytics kann man nun ganz einfach in Excel einer Maximalzeit gegenüberstellen, um zu ermitteln, wie viele Prozent des vorhandenen Textes im Schnitt überhaupt gelesen werden. Natürlich ist auch diese Methode nur sehr grob, weil am Ende niemand sagen kann, ob der Besucher abwesend auf die Seite starrt und an seinen letzten Urlaub denkt oder ob er wirklich liest. Hat man das Scroll-Tracking implementiert, kommt man auch aber hier der Wahrheit ein kleines Stück näher. Natürlich wird sich die URL-Lesedauer bezogen auf die Contentmenge in den seltensten Fällen 100 % nähern. Das tut der Nützlichkeit einer genaueren Messung aber keinen Abbruch. Je nach Domain und Inhalt werden sich bestimmte Prozentsätze einpendeln und es gilt wie immer, die positiven, vor allem aber die negativen Abweichungen von der individuellen „Norm“ zu filtern. Sieht man sich die Winner- und Loser-URL an, fallen einem sehr oft durch den neuen Fokus die Unterschiede ins Auge und man erkennt, was der (wahrscheinliche) Grund für Lesefreude oder -verweigerung ist. Dort setzt man dann mit Optimierungsmaßnahmen an.

Fleißaufgabe: Die Boilerplate ermitteln

Wie oben im letzten Hinweis erwähnt, ermittelt =WordCount die Gesamtmenge an Text einer Seite. Über „=CountWords“ hat man den Umfang des Primary Contents. Ermittelt man beides und zieht die Werte voneinander ab, bleibt die Boilerplate übrig, also die Anzahl Wörter, die auf allen oder strukturgleichen Seiten identisch oder fast identisch vorhanden ist. Der Wert wird sicherlich über einzelne Seiten etwas schwanken, aber über eine Mittelwertberechnung bekommt man dann schnell einen recht stabilen Kalkulationswert. Besteht die Boilerplate z. B. aus 850 Wörtern,

einem Wert, den man jetzt kennt, kann man diesen gedanklich im Umgang mit allen SEO-Tools bei Textauswertungen abziehen und landet dann beim Umfang des „echten“ Contents.

Bonustipp für Tracking-Nerds

Wer noch einen Schritt weiter gehen möchte, ermittelt in Excel aufgrund der errechneten maximalen Lesedauer Klassen wie z. B. 0.5, 1.0, 1.5 usw., um die Lesedauer größer zu kategorisieren. Diese Klassenbezeichnungen lassen sich dann auf URL-Basis als Tags über benutzerdefinierte Dimensionen und den Google-Tag-Manager in Analytics einspielen. Anschließend kann man dann dort nach Lesedauer oder auch nach (echter) Wortanzahl segmentieren.

Bonustipp für SEO

Aller Wahrscheinlichkeit nach und basierend auf einigen Bemerkungen von Mitarbeitern in einem anderen Zusammenhang ermittelt Google den Primary Content auf genau diese Weise. Man vergleicht die Strukturelemente einer Site via DOM und erkennt so ständig wiederkehrende Textelemente. Dort im DOM, wo nennenswerte Textmengen unikal, also von Seite zu Seite unterschiedlich sind, setzt man an und verwendet diese zur Textanalyse für das Ranking. Wer also seine eigentlich wichtigen Textinhalte je Seite kennt, ist bei Optimierungsmaßnahmen klar im Vorteil. Was Google mit seiner enormen Rechenpower zustande bringt, nämlich von unterschiedlich strukturierten Sites den Primary Content durch Mustervergleiche zu erkennen, müssen Sie eben für Ihre Domain manuell mit Gehirnschmalz, ein paar Mausclicks und Excel extrahieren. Wie das geht, wissen Sie ja jetzt. ¶