

CONSUMER



CUSTOMERS

Sebastian Erlhofer, Malte Müller, Tobias Häring

SEO-Data-Warehouse DAWIS

RESEARCH



Beinahe täglich stolpert man mittlerweile über das Buzzword Data-driven-Marketing. Gefühlt ist gerade alles und jeder Data-driven. Dennoch werden massiv Excel-Listen gepflegt, veraltete PHP-Skripte genutzt und die noch älteren Formelsammlungen und Makros als höheres Wissen gehandelt. Wie kann Data-driven-Marketing insbesondere im SEO aber eigentlich aussehen? Und wie kann ein SEO-Data-Warehouse hier helfen? Die SEO-Experten Sebastian Erlhofer, Malte Müller und Tobias Häring haben für die Beantwortung dieser Fragen umfassende Arbeit geleistet und das kostenlos nutzbare Open-Source-Projekt DAWIS aufgesetzt.



ANALYZING



DEMOGRAPHICS

BEHAVIOUR



DER AUTOR



Sebastian Erlhofer ist Geschäftsführer von mindshape, Buchautor und leidenschaftlicher Verfechter von gutem SEO, welches in der heutigen Zeit ohne Daten nicht mehr funktioniert.

DER AUTOR



Malte Müller ist SEO-Consultant bei mindshape in Köln. Am liebsten beschäftigt er sich mit Herausforderungen größerer Websites im Struktur- und Tech-SEO.

DER AUTOR



Tobias Häring ist Teamleiter der Data-PPC-Unit bei mindshape in Köln. In dieser Rolle fungiert er als Schnittstelle zwischen der hauseigenen Websolutions und der SEO-Unit.

STRATEGY

MARKETING

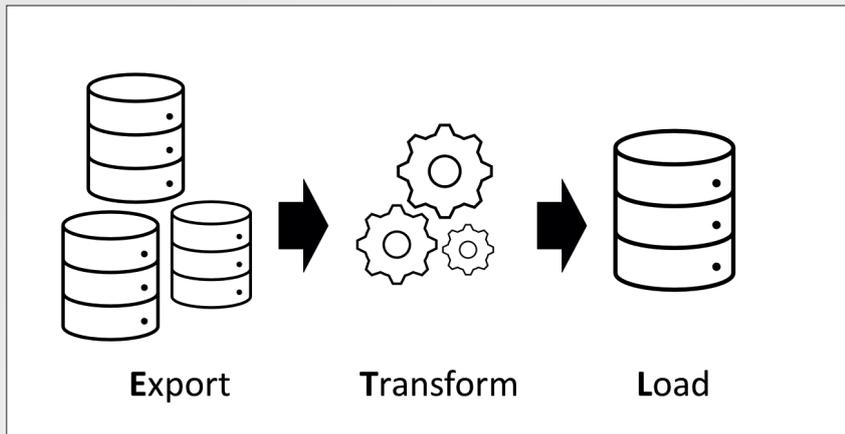


Abb. 1: Der ETL-Prozess überführt Daten strukturiert in ein Data-Warehouse

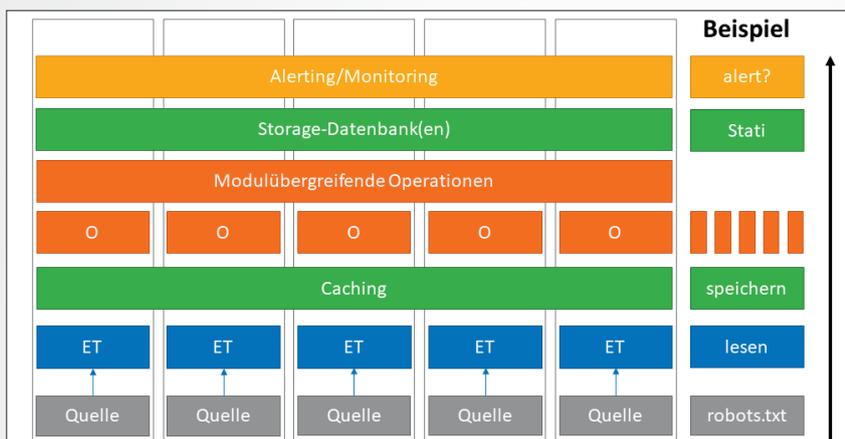


Abb. 2: Der Prozess eines Data-Warehouse mit DAWIS

Im zweiten Schritt werden die Daten dann transformiert. Das heißt, dass die im vorherigen Schritt erhobenen Daten umgewandelt, normiert und so angepasst werden, dass sie gut gespeichert werden können. Vor allem für unterschiedliche oder inkonsistente Datenquellen ist dies ein wichtiger Schritt, damit die spätere Verarbeitung auf gleichförmige Daten zugreifen kann.

Der dritte Schritt beinhaltet lediglich das Speichern der Daten in Datenbanken, um diese anschließend auswerten oder für Automatisierungen nutzen zu können. Hier ist vor allem eine technisch sichere und schnelle Speicherung ein Thema – und natürlich auch die schnelle Lesbarkeit für die folgenden Auswertungen.

Dieser ETL-Prozess speichert verschiedene Daten. Allerdings sind ja eigentlich die Auswertungen für SEOs spannend und wertschöpfend! Hier gab es lange Zeit nur eine wilde Mischung aus diversen Tools, Skripten und Excel-Vorlagen. Das Open-Source DAWIS („Data-Warehouse- und Information-System“ unter www.github.com/minds-hape-GmbH/dawis) geht einen Schritt weiter und bietet integrierte Verarbeitungs- und Auswertungsmöglichkeiten. Damit erweitert das System den ETL-Prozess um die zusätzlichen Ebenen „Operation“ und „Caching“.

Durch die zusätzlichen Ebenen im Prozess wird erreicht, dass zum einen auf Basis des Cachings bereits vor dem Abspeichern von Daten Prüfungen stattfinden können. So brauchen irrelevante Daten wie die Information, welche Seite zu welchem Zeitpunkt eine H1-Überschrift hatte, erst gar nicht im Data-Warehouse gespeichert werden, sondern es wird direkt eine Prüfung vorgenommen und nur dieses Ergebnis gespeichert. Die Caching-Ebene verbessert also den ETL-Prozess speziell für SEO-Anforderungen.

Wenn Sie sich die Klickraten (CTR) von Seiten anschauen und anschließend die Description oder den Title für die Google-Suchergebnisse anpassen, ist das eine datengetriebene Optimierungsmaßnahme, wenn auch eine kleine. Wie wäre es, wenn Ihnen regelmäßig und automatisch Vorschläge über nicht gut funktionierende URLs gemacht würden? Und zwar nicht nur auf Basis eines Faktors, der CTR, sondern auf Basis vieler verschiedener Datenpunkte aus unterschiedlichsten Quellen: zum Beispiel die CTR aus der Google Search Console, die Verweildauer aus Google Analytics und das Vorkommen des Hauptkeywords in der H1-Überschrift. Um solche Analysen automatisiert und regelmäßig durchführen zu können, sollten die benötigten Daten zentral an einem Punkt gespeichert und verfügbar gemacht werden: in einem Data-Warehouse.

Wie funktioniert ein Data-Warehouse?

Die Grundlage eines Data-Warehouse ist der sogenannte ETL-Prozess. ETL steht für Extract, Transform und Load und bildet die einzelnen Ebenen des Prozesses ab. In der ersten Ebene werden die Daten von einer Quelle extrahiert (extract). Welche Daten können das sein? Bei einem speziellen SEO-Data-Warehouse können das etwa Meta-Informationen der eigenen oder der Konkurrenz-Website sein. So können beispielsweise Title, Description oder andere Inhalte des HTML-Dokuments (H1-Überschrift, Indexierbarkeit usw.) erfasst werden. Aber auch Preisinformationen, strukturierte Daten oder andere Informationen sind denkbar. Als Datenquelle dienen neben der eigentlichen Website auch APIs, also die Schnittstellen von Tools, um beispielsweise die Zahlen der Google Search Console oder der SISTRIX-Toolbox nutzen zu können.

search_console

TABELLE ABFRAGEN TABELLE FREIGEBEN TABELLE KOPIEREN TABELLE LÖSCHEN EXPORT

Schema Details **Vorschau**

Zelle	property	date	clicks	impressions	ctr	position	searchType	clicksPreviousYear	clicksPreviousMonth	clicksPreviousWeek	clicksPreviousDay	impressionsPreviousYear	impressionsPreviousMon
1	https://www.mindshape.de/	2020-05-17	9.0	15.0		0.6	1.0 web	7.0	330.0	5.0	2.0	46.0	1695
2	https://www.mindshape.de/	2020-05-17	3.0	9.0	0.3333333333333331	1.8888888888888888	web	2.0	87.0	0.0	1.0	8.0	272
3	https://www.mindshape.de/	2020-05-17	3.0	25.0	0.12	4.04	web	0.0	19.0	0.0	0.0	1.0	1846
4	https://www.mindshape.de/	2020-05-17	1.0	15.0	0.06666666666666666	1.8666666666666667	web	1.0	25.0	1.0	1.0	29.0	1125
5	https://www.mindshape.de/	2020-05-17	1.0	13.0	0.076923076923076927		10.0 web	0.0	7.0	0.0	0.0	13.0	326
6	https://www.mindshape.de/	2020-05-17	1.0	20.0	0.05		11.85 web	null	2.0	0.0	1.0	null	996
7	https://www.mindshape.de/	2020-05-17	1.0	8.0		0.125	4.0 web	0.0	7.0	0.0	1.0	3.0	226
8	https://www.mindshape.de/	2020-05-17	1.0	7.0	0.14285714285714285		4.0 web	0.0	14.0	0.0	0.0	1.0	126
9	https://www.mindshape.de/	2020-05-17	0.0	6.0	0.0		72.0 web	null	0.0	0.0	0.0	null	255
10	https://www.mindshape.de/	2020-05-17	0.0	2.0	0.0		58.0 web	0.0	0.0	0.0	0.0	1.0	94
11	https://www.mindshape.de/	2020-05-17	0.0	2.0	0.0		55.5 web	0.0	0.0	0.0	0.0	1.0	106
12	https://www.mindshape.de/	2020-05-17	0.0	2.0	0.0		4.0 web	0.0	0.0	0.0	0.0	2.0	135
13	https://www.mindshape.de/	2020-05-17	0.0	8.0	0.0		50.125 web	0.0	2.0	0.0	0.0	1.0	124
14	https://www.mindshape.de/	2020-05-17	0.0	11.0	0.0	6.0909090909090908	web	0.0	0.0	0.0	0.0	11.0	305
15	https://www.mindshape.de/	2020-05-17	0.0	3.0	0.0		27.0 web	0.0	0.0	0.0	0.0	5.0	126

Abb. 3: Daten der Google Search Console in einer BigQuery

Entscheidend ist dann allerdings die Ebene der „Operation“. Hier finden die gewinnbringenden Auswertungen statt, die das Mehr aus den Daten herausholen und Daten miteinander in Verbindung setzen. Ein Beispiel ist das Clustern von Daten der GSC anhand von Brand-/ Non-Brand-Suchbegriffen oder in Bezug auf Seitentypen, was wiederum tiefergehende Auswertungen erlaubt. Damit ermöglicht DAWIS dann tiefere und detailliertere Auswertungen, die mit den Originalquellen gar nicht möglich sind – und das automatisiert ohne manuelles Zutun!

Technischer Aufbau

Das Warehouse DAWIS speichert Informationen in einer oder mehreren Datenbanken. Die Bandbreite reicht von lokalen SQL-Datenbanken bis hin zu Cloud-Datenbanken. Da beim Abspeichern der Rohdaten von Website-quellcodes oder aus der Google Search Console sehr viele Datensätze anfallen, lohnt sich ein Blick über selbstgehostete Datenbanken hinaus. Denn bei mehreren Millionen Zeilen geht eine übliche MySQL-Datenbank schnell in die Knie. Auch ist die Skalierbarkeit bei den cloudbasierten Datenbank-Systemen sehr einfach machbar. Auf dem Markt performanter Datenbanken für Data-Warehouses, die zum Abspeichern von Medium und Big Data geeignet sind, ist mittlerweile fast jeder der großen Online-Player vertreten. Häufig verwendete Plattformen sind Microsoft

Azure, Amazon Web Services (AWS) und Google Cloud Platform (GCP). Alle Anbieter haben neben den Datenbanken sehr umfangreiche weitere Services im Angebot, die im weiteren Verlauf und in Zukunft für ein SEO-Data-Warehouse interessant sein könnten. Dazu zählen beispielsweise Machine-Learning-Systeme. Die GCP hat den Vorteil, dass das kostenlose Dashboard- und Reporting-Tool Google DataStudio sehr gut mit dieser harmoniert und die sogenannte „BI-Engine“ dafür sorgt, dass sich selbst komplexe, umfangreiche Dashboards im Bruchteil einer Sekunde komplett aufbauen. Das ist mit dem direkten Connector zur Google-Search-Console-API nicht möglich. Und wenn man schon teils minutenlang auf eine Auswertung warten muss, wird man sie nicht häufig nutzen.

Die über DAWIS angebundene BigQuery-Datenbank lässt sich sehr einfach aufsetzen und skalieren. Natürlich sind solche Services nicht kostenlos, allerdings sind die anfallenden Kosten sehr überschaubar. Abgerechnet wird anhand zweier Faktoren: Speicherplatz und Traffic. Mit jeder Zeile Daten in der BigQuery steigt der benötigte Speicherplatz, welcher aber der wesentlich geringere Kostenfaktor bei der Berechnung ist. Die ersten 10 Gigabyte bekommen Sie geschenkt, jedes weitere Gigabyte kostet ca. zwei Cent pro Monat. So kommen Sie bei einem durchschnittlich großen Data-Warehouse mit einigen Dashboards und Alertings auf unter fünf Euro im

Monat. Natürlich ist ein wichtiger Punkt beim Speichern von Daten, dass die Vorgaben im Datenschutz eingehalten werden. So ist es möglich, den Serverstandort der Datenbank in der Europäischen Union zu wählen. Trotzdem sollten Sie sich bewusst machen, welche Daten Sie in einem Data-Warehouse speichern. Personenbezogene Informationen sind hier ein Problem und sollten am besten vermieden werden.

Modularer Aufbau

Das SEO-Data-Warehouse DAWIS ist modular aufgebaut. Geschrieben ist es in der Programmiersprache „Python“. Python ist nicht nur eine der meistgenutzten Programmiersprachen, wenn es um das Arbeiten mit Daten geht, sondern erlaubt durch die vergleichsweise einfache Syntax einen schnellen Einstieg in die Welt des Programmierens und das Arbeiten mit sehr großen Datenmengen. Durch den modulbasierten Aufbau von DAWIS wird ein flexibles System geschaffen, welches vergleichsweise einfach erweitert werden kann. Denn ein Modul erledigt immer nur eine Aufgabe. Derzeit sind folgende Module frei verfügbar:

- » **Robots.txt:** Steuert das Abrufen der robots.txt und das Überprüfen auf Veränderungen.
- » **HTML-Parser:** Fragt auf Basis einer URL-Liste HTML-Inhalte ab und überprüft deren Vorhandensein.
- » **PageSpeed Insights:** Auf Basis der URL-Liste sind Abfragen der Page-

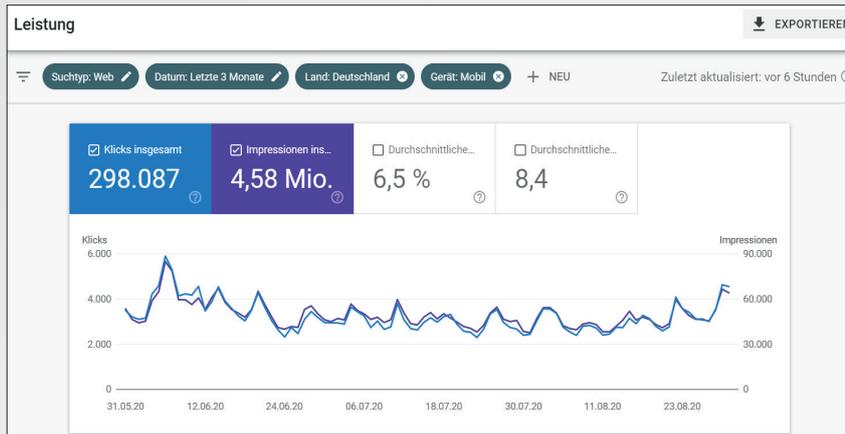


Abb. 4: Startseite des Leistungsberichts in der neuen Google Search Console

Speed-Insights-API von Google möglich.

- » **Google Search Console:** Abfragen der GSC-Search-API und Abspeichern der Daten.
- » **GSC Clustering:** Ordnet die GSC-Daten auf Basis selbst getroffener Filter einzelnen Clustern zu.
- » **SISTRIX Domain:** Fragt den Sichtbarkeitsindex der SISTRIX-Toolbox ab.

Nun sagen Sie vermutlich: „Moment. Abfragen von Inhalten einer Website und das Speichern der Daten der Google Search Console? Das ist doch nichts Neues.“ Stimmt. Es gibt auf dem Markt sehr viele Tools, die hervorragende Arbeit leisten und die jeweils einzelne der genannten Aufgaben beherrschen. Nehmen wir zum Beispiel Audisto mit dem neuen Monitoring-Tool (www.audisto.com/monitoring). Ein weiteres Tool mit ähnlichen Möglichkeiten ist Testomato (www.testomato.com). Auch hier sind Website-Checks möglich, sodass Sie Indexierungsvorgaben, Überschriften und die Robots.txt überprüfen sowie Ist-Soll-Werte abgleichen können. Für die regelmäßige Abfrage der API der Google Search Console und das Abspeichern der Daten in der Google BigQuery hingegen gibt es beispielsweise Supermetrics (www.supermetrics.com/product/bigquery).

Diese und andere Tools decken meist gut die speziellen Einsatzzwecke ab. Zwei Dinge sprechen aber für ein Data-Warehouse im Gegensatz zu Einzellösungen: Zum einen sind die Ein-

satzzwecke meist Insellösungen (und wollen auch einzeln bezahlt werden). Das heißt, der zugrunde liegende Zweck wird sehr gut erfüllt. Ein Blick über den Tellerrand der Tools hinaus ist aber meist nicht möglich. Zum anderen sind Insellösungen schwer miteinander in Verbindung zu setzen, eine Aggregation der Daten ist quasi unmöglich oder manuell zumindest sehr aufwendig. Erst mit einem zentralen Data-Warehouse können Sie einzelne Datensätze aus unterschiedlichsten Quellen miteinander verknüpfen und kombinieren und so völlig neue und gewinnbringende Auswertungsmöglichkeiten schaffen. Das kann im Vergleich zum Wettbewerb einen entscheidenden Vorteil bieten!

Sind die Daten erst einmal zentral gespeichert, stehen zudem auch zahlreiche allgemeine und individuelle Auswertungsmöglichkeiten zur Verfügung: Ob Google DataStudio, Grafana, Power BI, Tableau oder Python direkt – alles ist möglich. Selbst das von vielen SEOs heiß geliebte Excel kann in Kombination mit Excel-Add-Ins direkt auf die DAWIS-Datenbanken zugreifen.

SEO-Data-Warehouse im Praxiseinsatz

Doch genug der Theorie: Wie können Ihnen die Daten aus einem Data-Warehouse jetzt konkret im SEO-Alltag helfen? Dazu folgen drei Bereiche, in denen Ihnen ein SEO-Data-Warehouse in der täglichen Arbeit zu mehr Effizienz oder Qualität verhilft:

INFO

RegEx sind Zeichenfolgen, die einer bestimmten Syntax folgen, um wiederum andere Zeichenketten auf ein oder mehrere Kriterien zu untersuchen. Der RegEx „`^test[1-3]!`“ gibt z. B. vor, dass die untersuchte Zeichenkette mit dem String „test“ beginnen, darauf eine Zahl zwischen eins und drei folgen und am Ende ein Ausrufezeichen stehen muss. In der Google Search Console könnten RegEx genutzt werden, um Suchanfragen oder URLs zu clustern. Eine Übersicht zur Syntax, die Google verwendet, finden Sie in der Google-Analytics-Hilfe unter folgendem Link: <http://einfach.st/regex>.

1. Google Search Console + DataStudio
2. Onpage-Monitoring
3. Verknüpfung von Daten

Google Search Console + DataStudio = ❤️

Die Daten aus der Google Search Console sind die besten, die Google für die Arbeit im SEO zur Verfügung stellt. Daran ändern auch die vielen Inkonsistenzen nichts, die vor allem seit der Umstellung auf die neue Google Search Console immer häufiger auftreten. An keiner anderen Stelle erhalten SEOs so direkten Zugriff auf die Leistungsdaten der eigenen Website in der Suche. Das Zusammenspiel aus Klicks, Impressionen und Position bietet mit den richtigen Auswertungen eine wertvolle Grundlage, um Probleme und Potenziale zu erkennen.

Insbesondere bei großen Websites und komplexen Auswertungen gerät die Benutzeroberfläche des Leistungsbereichs der Google Search Console im Browser allerdings schnell an ihre Grenzen. Dafür sind verschiedene Faktoren verantwortlich:

- » Pro Dimension ist nur ein Filter möglich: Sie können also nicht gleichzeitig auf Suchanfragen mit „Sofa“ oder „Sessel“ filtern. Das erschwert je nach Fragestellung die Analyse enorm.

- » Die maximale Zeilenanzahl in den Tabellen und beim Export beträgt 1.000 Zeilen: Bei einer Website mit 100.000 URLs können Sie demnach nur einen Bruchteil der Daten exportieren. Möchten Sie die Daten aus der Google Search Console also im großen Stil analysieren oder weiterverarbeiten, stehen Sie vor einem Problem. Auch bei kleineren Websites kann zudem der Bericht zu den Suchanfragen schnell die maximale Zeilenzahl überschreiten.
- » Die Verwendung regulärer Ausdrücke wird (noch) nicht unterstützt: Suchanfragen oder URLs können nicht auf Basis von RegEx gefiltert oder geclustert werden. Das macht viele Auswertungen wie den Vergleich von Seitentypen oder gleichartigen Suchanfragen unmöglich.

Für Data-driven-SEO ist die Nutzung der Google-Search-Console-API also unumgänglich. In Kombination mit dem Weiterverarbeiten der Rohdaten und anschließendem Aufbereiten in einem Analyse-Tool Ihrer Wahl werden dann auch viele bis dato undenkbar Auswertungen möglich. Im Folgenden möchten wir vier Bereiche kurz ins Auge fassen: Monitoring, Reporting, Tiefenanalysen und hypothesenbasiertes Testing. Los gehts.

Monitoring

Ein wichtiger Arbeitsbereich im SEO besteht seit jeher im Monitoring. Ranking- und Traffic-Entwicklungen der eigenen Website müssen regelmäßig überwacht werden, um auf Schwankungen frühestmöglich reagieren zu können. Leider findet dieses Monitoring oft zu oberflächlich statt: Bei einer Website mit verschiedenen Seitentypen und Suchintentionen macht es keinen Sinn, sich einfach nur die Gesamtentwicklung der Klicks anzuschauen. Zu unterschiedlich sind die Inhalte auf den eigenen Seiten und

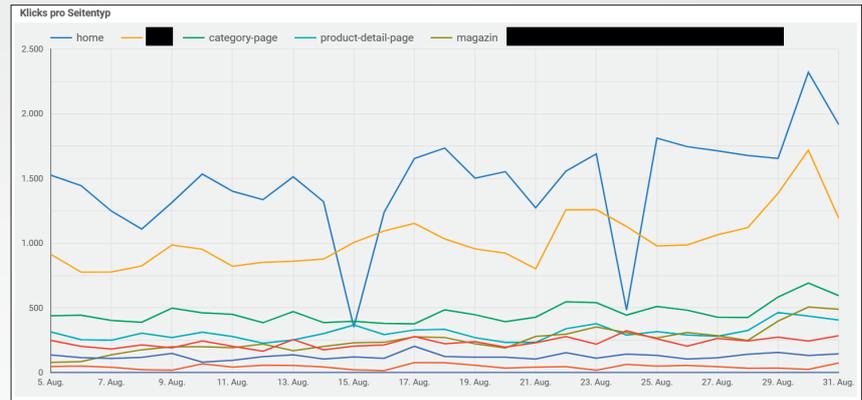


Abb. 5: Klicks nach Seitentyp im Verlauf

URL-Struktur	Beispiel
www.domain.de/silo1/	Backwaren
www.domain.de/silo1/unterthema1/	(Backwaren) – Sahnetorten
www.domain.de/silo1/unterthema1/produkt1/	(Backwaren – Sahnetorten) – Schwarzwälder-Kirschen
www.domain.de/silo1/unterthema1/produkt2/	(Backwaren – Sahnetorten) – Bienenstich
www.domain.de/silo1/unterthema2/	(Backwaren – Sahnetorten) – Kuchen
www.domain.de/silo1/unterthema3/	(Backwaren – Sahnetorten) – Kleingebäck
www.domain.de/silo2/	Käsespezialitäten
www.domain.de/silo3/	Getränkesspezialitäten

Abb. 6: URL-Struktur mit Siloing

ebenso die SERPs für die verschiedenen Suchmotivationen. Daher ist eine ordentliche Segmentierung wichtig. Oft bietet sich ein Aggregieren bzw. Clustern der GSC-Daten nach Seitentypen an, wie es beispielhaft für einen Online-Shop in Abbildung 5 zu sehen ist.

Das Beispiel zeigt eine erste Segmentierung nach Seitentypen. Je nach Website kann diese noch deutlich komplexer ausfallen. So können weitere Seitentypen wie spezielle Landingpages hinzukommen oder innerhalb von Seitentypen nochmals unterschieden werden. Beispielsweise kann es Sinn machen, Kategorie-seiten in einem Online-Shop nach der Ebene zu clustern, in der sie in der Website-Struktur aufgehängt und verlinkt sind. Eine prominent platzierte Kategorie-seite zu „Esszimmermöbeln“ kann sich anders verhalten als eine sehr spezifische Kategorie-seite zu „runden Esstischen“. Nicht nur die dahinterliegenden Suchanfragen sind deutlich spezifischer, auch intern sind die Seiten ganz unterschiedlich verlinkt. In diesem Fall

würde man den Seitentyp der Kategorie-seiten nochmals feiner segmentieren und Metriken wie Rankings oder die Klickrate getrennt überwachen.

Neben dem Clustern nach Seitentyp bietet sich auch häufig die Betrachtung einzelner Verzeichnisse im Vergleich an. Voraussetzung dafür ist eine sprechende und hierarchische URL-Struktur, in der thematisch zusammenhängende Bereiche in Form von Verzeichnissen gruppiert werden. Ihnen wird diese Technik sicherlich auch als „Siloing“ bekannt sein. Den möglichen Aufbau einer solchen URL-Struktur sehen Sie in Abbildung 6. Durch verschiedene Faktoren ist es gut möglich, dass Ihre Website in unterschiedlichen thematischen Bereichen auch unterschiedlich gut performt. Gilt Ihre Website als Autorität im Bereich „Backwaren“, erreichen Sie dort wahrscheinlich leichter gute Rankings als im Bereich „Getränkesspezialitäten“. Ebenso kann sich je nach Thema die Konkurrenz-situation stark unterscheiden, da Ihre SERPs-Wettbewerber stärker oder schwächer ausfallen.

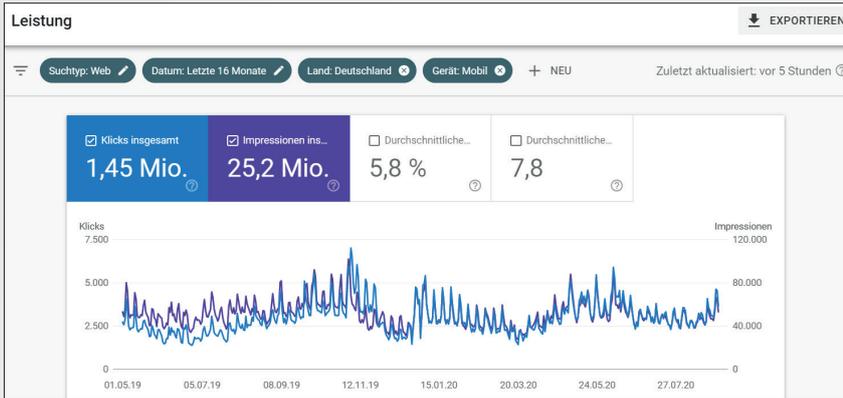


Abb. 7: Bei langen Zeiträumen ist die Aggregation auf Tagesdimension wenig aussagekräftig

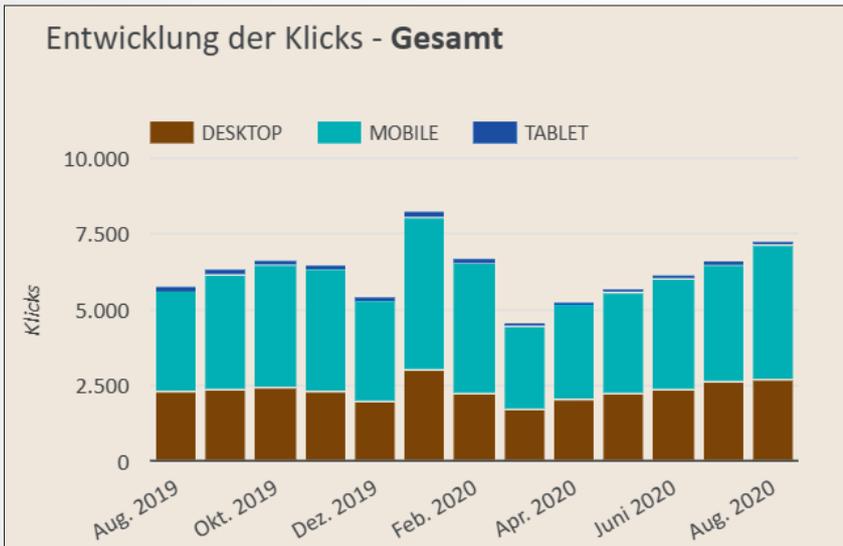


Abb. 8: Die Aggregation auf Monatsdimension macht für lange Zeiträume deutlich mehr Sinn

Wichtige Seitentypen und Verzeichnisse können oft regelbasiert über die URL-Struktur geclustert werden. Hier wird die Bedeutung der DAWIS-„Operation“-Ebene deutlich. Daneben gibt es aber mit Sicherheit auch Cluster, die Sie nur manuell zusammenstellen können. Im E-Commerce könnten das z. B. bestimmte Fokus-Sortimente sein, mit denen Sie besonders hohe Margen erzielen. Da diese aber über die URL-Struktur nicht eindeutig identifizierbar sind, müssen nach dem Aggregieren der einzelnen Datenpunkte manuelle Dimensionen hinzugefügt werden können. Dies geschieht in DAWIS beispielsweise auf Basis eines CSV-Imports und anschließendem Abgleich. So werden für jede URL die Informationen hinzugefügt, ob sie Teil des Fokus-Sortiments ist oder nicht. Damit können Sie dann anschließend eigene Auswertun-

gen für Ihre Fokus-Sortimente erstellen.

Achten Sie im Monitoring also auf zielführende Segmente wie Seitentypen und Verzeichnisse. Auf dieser Datengrundlage können Sie datengetrieben deutlich tiefgründigere Schlüsse ziehen und Entscheidungen treffen.

Reporting

Die eigene Speicherung und Weiterverarbeitung von Google-Search-Console-Daten hilft Ihnen nicht nur im Monitoring, sondern ermöglicht auch neue Chancen im Reporting.

Die letzten 16 Monate sind der maximale Zeitraum, der Ihnen in der Arbeit mit der Google Search Console zur Verfügung steht. Daran ändert auch die Abfrage der API nichts. Ältere Daten sind nicht verfügbar, was im Reporting zum Problem werden kann, wenn Sie längere Zeiträume vergleichen wollen. Stellen Sie sich vor, Ihr Vorgesetzter bittet Sie zum Jahresbeginn um eine Präsentation zur Entwicklung des SEO-Kanals im vergangenen Jahr. Um Ihren Erfolg aufzuzeigen, möchten Sie die Entwicklung der Non-Brand-Klicks pro Monat im Vergleich zum Vorjahr präsentieren. Leider wird das nicht funktionieren: Auf die Daten für das erste Halbjahr des Vorjahres können Sie in der Google Search Console nicht mehr zugreifen. Sie sehen: Es ist auch bei 16 Monaten wichtig, die Leistungsdaten aus der Google Search Console regelmäßig abzufragen und selbst zu speichern.

Dazu kommen die Einschränkungen bei der zeitlichen Aggregation: In der Weboberfläche steht Ihnen ausschließlich die Tagesdimension zur Verfügung. Das macht für die Auswertung längerer Zeiträume meist keinen Sinn. Durch die

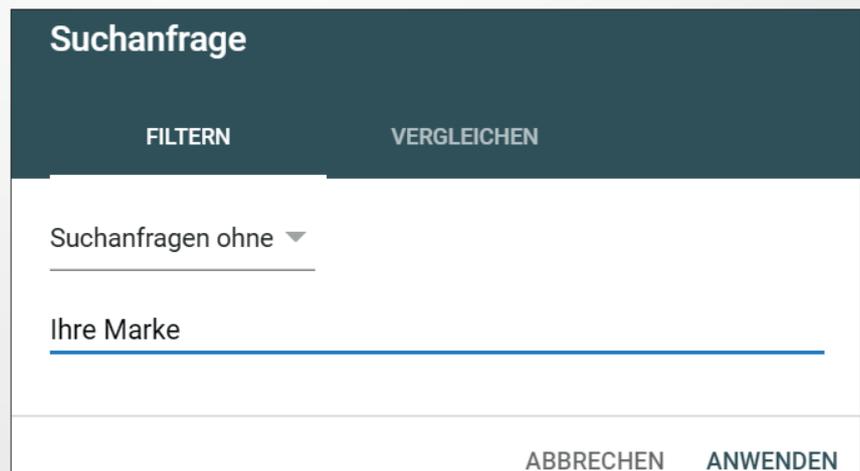


Abb. 9: Eine Non-Brand-Filterung in der Google Search Console ist oft nicht genau genug

Query	Klicks vorletzte 7 Tage	Klicks letzte 7 Tage	Klicks Δ ^	Klicks Δ %	Position Δ	Position letzte 7 Tage
1.	20	null	null	null	null	null
2.	3.135	2.941	-194	-6,19 %	-0,15	2,51
3.	902	795	-107	-11,86 %	-0,13	1,46
4.	1.127	1.033	-94	-8,34 %	0,01	1,04
5.	253	163	-90	-35,57 %	0,55	6,35
6.	853	768	-85	-9,96 %	0,17	1,15
7.	261	182	-79	-30,27 %	-1,09	7,7
8.	416	340	-76	-18,27 %	0	1
9.	82	7	-75	-91,46 %	0,21	1,8
10.	71	0	-71	-100 %	-10,85	12,18
11.	377	311	-66	-17,51 %	-0,41	6,57
12.	303	240	-63	-20,79 %	-0,47	5,03
13.	378	318	-60	-15,87 %	-0,43	3,21
14.	188	134	-54	-28,72 %	-2,01	3,28

Abb. 10: Durch die Sortierung nach der Klick-Differenz zwischen zwei Zeiträumen sind die größten Verlierer schnell zu erkennen

zu granulare zeitliche Dimension entsteht in den meisten Fällen ein Graph im Zickzack-Muster (siehe Abbildung 7), der sich nur bedingt interpretieren lässt. Eine Auswahl zwischen der Aggregation auf Tages-, Wochen- und Monatsbasis, wie sie z. B. in Google Analytics zur Verfügung steht, existiert in der Google Search Console nicht.

Abhilfe schafft auch hier eine Einstellung im Google DataStudio. Duplizieren Sie dafür die Dimension „Date“ und stellen Sie die Kopie auf „Jahr und Monat“ ein. Wählen Sie dazu als Diagrammtyp ein Säulendiagramm und schlüsseln zusätzlich noch nach Geräteklassen auf, erhalten Sie eine Ansicht wie in Abbildung 8. Noch freier sind Sie, wenn Sie in Ihrem Data-Warehouse direkt verschiedene zeitliche Aggregationen speichern. In DAWIS laufen diese zusätzlichen Abfragen als automatische Operationen. So können Sie beispielsweise all Ihre Abfragen nicht nur auf Tagesdimension stattfinden lassen, sondern zusätzlich auch regelmäßige Abfragen auf Kalenderwochen und Monate durchführen. Dadurch muss die Neuberechnung nicht erst im DataStudio stattfinden, was je nach Datenmenge die Ladezeit erheblich verkürzt.

Nicht nur neue Möglichkeiten bei Zeiträumen und Zeitdimensionen sind spannend: Auch die eben bereits erwähnte Unterscheidung in Brand- und Non-Brand-Segmente ist für das Reporting essenziell. Bei der Bewertung des

organischen Traffics sollten Sie für die meisten Fragestellungen alle Suchanfragen ausschließen, die sich um die eigene Marke drehen. Bei Brand-Suchanfragen zeigen sich gänzlich andere Nutzersignale, welche die Gesamtwerte enorm verzerren können. Ein Beispiel dafür sind unverhältnismäßig hohe CTR-Werte, da die Nutzer Ihre Marke schon kennen und sich bewusst für Ihr Ergebnis entscheiden werden.

Also müssen Brand- und Non-Brand-Suchanfragen voneinander getrennt (segmentiert) werden. Aktivieren Sie dafür versuchsweise einmal in der GSC-Weboberfläche den Filter für die Suchanfragen und tragen Sie unter der Option „Suchanfragen ohne“ Ihren Markennamen ein (siehe Abbildung 9).

Was sich erst einmal einfach anhört, erweist sich in der Praxis leider oft als ungenügend. Denn es gibt viele Möglichkeiten, warum eigentliche Brand-Suchanfragen durch einen solch einfachen Filter nicht erfasst werden. Häufige Beispiele dafür sind Falschschreibweisen oder Eigennamen hauseigener Produkte oder Dienstleistungen, die ebenfalls dem Brand-Bereich zuzuordnen sind. Anstatt des einfachen IN-Filters in der Weboberfläche benötigen Sie also eine Segmentierung der Suchanfragen per RegEx, welche die GSC derzeit nicht bietet. In DAWIS erfolgt diese Segmentierung als nachgelagerte Operation auf die Daten – flexibel und vor allem automatisch.

HINWEIS

Wie immer bei der Arbeit mit Google-Search-Console-Daten gilt: Filtern Sie möglichst genau, um zu stark aggregierte Durchschnittswerte zu vermeiden. Das gilt vor allem bei der Arbeit mit CTR- oder Positions-Daten. Der Filter auf Ihr Zielland ist beispielsweise Pflicht, um schlechte Zufallsrankings in anderen Ländern auszuschließen. Diese verzerren Ihre durchschnittlichen Positionen ansonsten extrem.

Tiefenanalyse

Kennen Sie das? Es ist Montagmorgen, Sie loggen sich bei SISTRIX ein und bekommen fast einen Herzstillstand: Nach dem vor Kurzem ausgerollten Core-Update ist Ihre Sichtbarkeit um 40 Prozent eingebrochen. Wie immer gilt: Keine Panik! Nach dem ersten Kaffee wird es also Zeit, sich die Entwicklung genauer anzuschauen. Früher mussten Sie erst aufwendig Daten ziehen, Excel-Tabellen mit Pivot-Tabellen bemühen und und und ... Zum Glück haben Sie heute ein SEO-Data-Warehouse mit segmentiertem Monitoring auf Google-Search-Console-Daten, sodass Sie in Ihren Dashboards schnell die ersten Antworten finden auf Fragen wie:

- » Wie haben sich meine Leistungsdaten in der Suche entwickelt?
- » Sind meine Fokus-URLs betroffen?
- » Zeigen sich Auffälligkeiten zwischen den einzelnen Seitentypen?
- » Zeigen sich Auffälligkeiten zwischen den einzelnen Verzeichnissen?

Ist der Großteil Ihrer Verzeichnisse konstant geblieben und nur eines zeigt einen deutlichen Abfall, wissen Sie, wo sie im nächsten Schritt suchen müssen. Jetzt wird es Zeit für die Tiefenanalyse. Dazu nutzen Sie Ihre angereicherten BigQuery-Daten mithilfe individueller Dashboards oder des Explorers in Google DataStudio. Ebenso lassen sich diese gezielt exportieren, um anschließend Analysen in Excel und Co. durchzuführen. Es gilt herauszufinden, was genau in diesem Verzeichnis passiert ist. Filtern Sie in Ihren Google-Search-Console-Daten also auf dieses Verzeichnis und das Non-Brand-Segment – zum Analysieren von Update-Veränderungen sind Brand-Suchanfragen meist auszuschließen, da diese nur das Gesamtbild verzerren. Bauen Sie sich anschließend Auswertungen für Fragen wie:

- » **Welche URLs haben am meisten Klicks verloren?** Standardmäßig werden URLs oft absteigend nach ihren absoluten Klicks sortiert. Um Veränderungen zu erkennen, sortieren Sie die URLs besser absteigend nach ihrer Klickdifferenz im untersuchten Zeitraum zum vorherigen Zeitraum bzw. Vorjahreszeitraum. So sehen Sie auf einen Blick, wo die größten Verluste stattgefunden haben. Was sich einfach anhört, gestaltet sich in der Praxis oft schwierig: Tools wie das Google DataStudio unterstützen zwar automatische Berechnungen für Vergleichszeiträume, können aber nicht nach der Differenz sortieren. Abhilfe schafft eine gesonderte Operation in Ihrem Data-Warehouse.
- » **Welche Non-Brand-Keywords haben am meisten Klicks verloren?** Traffic-relevante Ranking-Verluste auf der ersten SERPs zeigen sich meist in deutlichen Klickrückgängen. Um ein geringeres Suchvolumen für das jeweilige Keyword im untersuchten Zeitraum auszuschließen, eignet

sich eine Validierung mit der CTR- oder Positionsveränderung. Eine mögliche Aufbereitung sehen Sie in Abbildung 10. Dadurch, dass zur Differenz in den Klicks auch die Differenz in der Position enthalten ist, können Sie erkennen, ob mit dem Klickverlust auch ein Positionsverlust einhergeht. In welchem Bereich der SERPs sich der Positionsverlust ungefähr abgespielt hat, erkennen Sie über die durchschnittliche Position der letzten sieben Tage.

- » **Welche Non-Brand-Keywords haben am meisten Impressions verloren?** Rankingverluste wichtiger Keywords auf die zweite Seite zeigen sich am ehesten über deutliche Impressionsrückgänge. Auch hier können Sie die Entwicklung über die Positionsveränderung validieren.
- » **Haben die Verluste eher im Short-head, Midtail oder Longtail stattgefunden?** Dafür bietet sich eine Auswertung an, welche den Anteil an den Non-Brand-Klicks je Länge der Suchanfrage beinhaltet.
- » **Wie hat sich die Anzahl an Suchanfragen mit mindestens X Klicks (pro URL) verändert?** Wer schon mal mit Google-Search-Console-Daten gearbeitet hat, weiß, wie wichtig Schwellenwerte sein können. Sie dienen der Bereinigung und Fokussierung auf die wesentlichen Daten. Rankingverluste zu Keywords, über die ohnehin keine Klicks generiert wurden, sind meist irrelevant.

Sie sehen: Die Arbeit mit Google-Search-Console-Daten muss keinesfalls beim Monitoring enden. Auch für viele Tiefenanalysen sind sie eine ideale Arbeitsgrundlage, sobald sie einmal strukturiert und segmentiert in einem Data-Warehouse zur Verfügung stehen. Durch das eigene Verwalten der Daten sind Sie zudem nicht an die vorgefertigten Berichte gängiger Tools gebunden, sondern können passgenaue

Auswertungen für Ihre individuellen Fragestellungen entwerfen.

Hypothesenbasiertes Testing

Data-Driven-SEO bedeutet, dass Entscheidungen aufgrund der Analyse und Interpretation von Daten getroffen werden. Sie lesen also in Ihrer Lieblings-Fachzeitschrift Website Boosting einen Artikel zum Thema FAQs und strukturierte Daten. Der Autor berichtet von CTR-Verbesserungen und bringt Sie damit auf die Idee, für Ihre Seiten ebenfalls über die Erstellung und Auszeichnung von FAQs nachzudenken. Das Problem ist nur: Sie benötigen Freigabe „von oben“ für die Content-Ressourcen, um die FAQs zu produzieren. Gleichermaßen brauchen Sie Unterstützung aus der IT für das Einpflegen der strukturierten Daten. Wie überzeugen Sie also Ihren Vorgesetzten von dieser Idee? Genau, mit Daten!

Bevor Sie nach Ressourcen für einige Hundert Seiten fragen, erbitten Sie erst mal die nötige Unterstützung für wenige URLs. Implementieren Sie nur auf diesen die FAQs inkl. Auszeichnung und fügen Sie sie mithilfe Ihres Data-Warehouse zu einem eigenen Cluster zusammen. So können Sie sich ein Test-Set kreieren, welches Sie ab dann gesondert beobachten können – ein etabliertes Vorgehen aus der Conversion-Optimierung. Sie testen also eine Vermutung zunächst im kleinen und günstigen Rahmen, bevor Sie gleich eine vermeintliche Optimierung überall ausrollen. Dieses Vorgehen bezeichnet man als hypothesenbasiertes Testing. Das hat gleich mehrere Vorteile:

1. Sie testen für Ihre Website und Ihr SERPs-Umfeld aus, ob die Haupt-Keywords Ihrer URLs im Test-Set positive CTR-Veränderungen aufweisen. Was für andere Websites gut klappt, muss für Ihre nicht ebenfalls gut funktionieren.

2. Durch die bereits existierenden Filter- und Cluster-Möglichkeiten im Data-Warehouse können Sie sich im Google DataStudio in wenigen Schritten ein Dashboard bauen, welches die CTR-Entwicklung beinhaltet. Das spart nicht nur Ihnen bei der Erfolgsbewertung der Maßnahme Zeit, sondern kann auch super als Kommunikationsinstrument gegenüber Ihrem Vorgesetzten genutzt werden.
3. Die Verknüpfung von Daten erlaubt Ihnen, verschiedene relevante Datensätze in Relation zu setzen. Im Falle des FAQ-Markups wären z. B. die Informationen wichtig, ob die entsprechenden URLs bereits gecrawlt wurden oder für welche URLs das Markup überhaupt zuverlässig ausgespielt wird. Wird das Markup schließlich nur teilweise ausgespielt und Sie schließen die betroffenen URLs nicht aus, kann das Ihre Auswertung verzerren. Mehr zu der Verknüpfung von Daten und den daraus resultierenden Möglichkeiten erfahren Sie gegen Ende des Artikels.

Haben Sie den Erfolg der Maßnahme einmal transparent gemacht, haben Sie das perfekte Argument, um die Freigabe für die restlichen URLs zu bekommen. Und ganz nebenbei haben Sie auch noch deutlich effizienter gearbeitet und Ressourcen geschont, wenn die Tests nicht wie erhofft gelaufen sind.

Das Szenario des FAQ-Markups ist nur eins von vielen Beispielen, in denen sich ein solches Vorgehen anbietet. Grundsätzlich sind derartige Auswertungen immer spannend, um eigene Hypothesen zu testen. Andere Ideen wären z. B. die Hypothesen-getriebene Optimierung von Snippets oder der Vergleich von Kategorienseiten mit und ohne Kaufberater. In beiden Fällen bieten die Google-Search-Console-Daten einen guten Ansatzpunkt. Auch wenn sich im SEO durch die vielen unbekannt Faktoren keine statistisch exakten

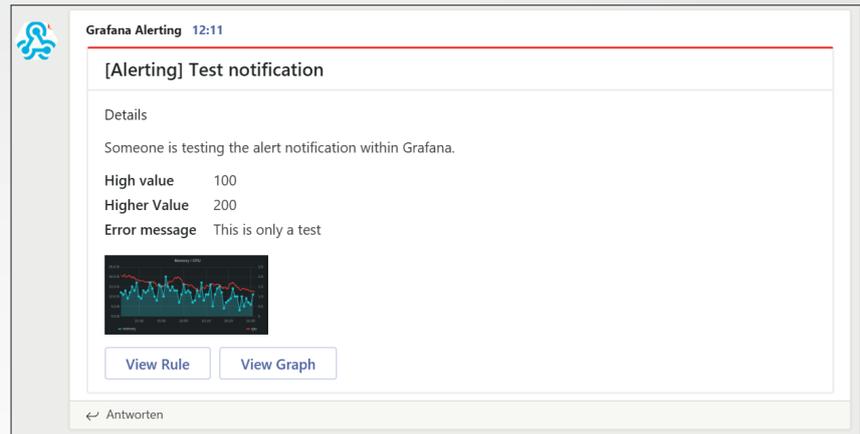


Abb. 11: Lassen Sie sich über Abweichungen direkt in Microsoft Teams benachrichtigen.

Tests durchführen lassen, bietet sich vor allen in komplexen Organisationsstrukturen ein solches Vorgehen sehr an.

Onpage-Monitoring

Die Daten eines SEO-Data-Warehouse sind nicht nur auf die Suchdaten aus der Google Search Console beschränkt. Auch Onpage-Daten von URLs spielen eine wichtige Rolle, welche in DAWIS primär mit einem HTML-Parser erhoben werden. Stellen Sie sich vor, dass bei einem Website-Deployment versehentlich die Titles aller Kategorienseiten Ihres Online-Shops verloren gehen. Wenn Sie nicht zufällig darauf stoßen, bekommen Sie den Bug wahrscheinlich erst mit, wenn er sich im Traffic niederschlägt. Das ist zu spät. Die Lösung ist einfach und heißt automatisches Onpage-Monitoring. Als Vorbereitung hierzu müssen Sie sich überlegen, welche Elemente Ihrer Website kritisch für den SEO-Erfolg sind. Während es immer auch einige projektspezifische Kriterien gibt, existieren viele pauschale Prüfungen, die auf jeder Website Sinn machen. Dazu zählen beispielsweise:

- » Statuscodes: Sind wichtige Seiten noch mit dem Statuscode 200 erreichbar?
- » Meta Robots: Sind wichtige Seiten noch indexierbar?
- » Meta Title: Haben wichtige Seiten noch einen Title? Stimmen Soll- und Ist-Title noch überein?

- » H1: Haben wichtige Seiten noch eine Hauptüberschrift? Stimmen Soll- und Ist-Überschrift noch überein?
- » Strukturierte Daten: Enthalten wichtige Seiten noch die strukturierten Daten?

Diese Liste lässt sich noch lange fortsetzen. Je nach Art der Website muss sich das Monitoring nicht auf SEO-Onpage-Faktoren beschränken. Einen Klassiker im E-Commerce stellt z. B. die Prüfung des Preises auf Produktdetailseiten > 0 Euro dar. Spielt der Log-in-Bereich auf einer Website eine wichtige Rolle, kann das Vorhandensein des „Anmelden“-Buttons geprüft werden. Dabei können die Prüfungen auf verschiedenen Ebenen stattfinden. Oft reicht es nicht aus, zu wissen, dass ein Element vorhanden ist. Stattdessen kann es auch wichtig sein, wie oft ein Element vorhanden ist oder welchen Wert das Element aufweist. Beispielsweise wollen Sie keine mehrfachen Canonicals auf einer URL oder müssen sich sicher sein, dass sich der Inhalt des Titles gegenüber dem Vortag nicht geändert hat.

Vor der Implementierung stellt sich noch die Frage, welche und wie viele URLs im Monitoring berücksichtigt werden sollen. Während bei kleinen Websites die Überwachung aller URLs noch denkbar ist, stellt das bei großen Domains mit mehreren 100.000 Unterseiten keine Option dar. Stattdessen bietet sich dort ein Template-basiertes Monitoring an. In dieser

URL mit 404-Statuscode	Zugriffe durch Googlebot	Seitenaufrufe in Analytics
1.	125	null
2.	87	352
3.	84	16.720
4.	56	172
5.	55	965
6.	51	329
7.	51	526
8.	50	320
9.	49	4.641
10.	47	522
11.	47	619
12.	45	136
13.	45	236

Abb. 12: Die Verbindung aus Logfile-Daten des Googlebots und 404-Tracking in Analytics bietet eine gute Grundlage zur Bewertung der Wichtigkeit von 404-Fehlern

Größenordnung finden schließlich nur selten manuelle Anpassungen an einzelnen URLs statt, meist wirken sich Änderungen auf alle Seiten des gleichen Seitentyps aus. Eine Zusammenstellung der wichtigsten Templates für einen Online-Shop könnte z. B. folgendermaßen aussehen:

- » Startseite
- » Kategorienseiten
- » Produktdetailseiten
- » Themenseiten
- » Suchergebnisseiten
- » Magazinseiten
- » Check-out
- » Serviceseiten (Impressum, Kontakt & Co.)

Anstatt also alle URLs zu überwachen, beziehen Sie nur einige repräsentative URLs jedes Seitentyps in das Monitoring ein. In DAWIS erfolgt die Auswahl der Seiten auf Basis von URL-Sets, die Sie in der YAML-Konfigurationsdatei anlegen und pflegen können. Jedes Template könnte in diesem Fall einem URL-Set entsprechen. So können Sie auch die gewünschten Prüfungen auf die jeweiligen Templates abstimmen: Während auf Kategorienseiten die Kontrolle des Titles besonders wichtig sein kann, ist auf internen Suchergebnisseiten vielleicht das Vorhandensein des Noindex-Tags am relevantesten. Definieren Sie für die einzelnen Seitentypen, welche Kriterien Sie jeweils überwachen wollen. Im gleichen Zuge sollten Sie sich über-

legen, in welchem Intervall die einzelnen Prüfungen erfolgen sollen. Für den Statuscode der Startseite beispielsweise empfiehlt sich ein sehr kurzes Intervall von fünfminütlichen Checks. Für die Kontrolle der Breadcrumb-Auszeichnung auf Produktdetailseiten hingegen genügt ein längeres Intervall von z. B. täglichen Checks. Je größer und direkter die potenziellen Auswirkungen des Tests sind, desto öfter muss das Kriterium geprüft werden und desto dringlicher müssen Sie im Fall der Fälle handeln.

Neben dem Onpage-Monitoring von Templates kann je nach Website-Größe und -Art auch ein regelmäßiges Monitoring durch komplette Crawls sinnvoll sein. Dabei steht weniger die strukturierte Kontrolle einzelner Seitenelemente im Mittelpunkt, als sonstige Auffälligkeiten durch eher explorative Abgleiche zu erkennen. Dazu zählen beispielsweise Veränderungen an der Gesamtzahl an URLs oder im Vergleich zum vorherigen Crawl neu oder nicht mehr gefundene URLs. In DAWIS ist diese Funktion aufgrund der Veröffentlichung als Open Source und damit verbundener Hürden für ein Crawling nicht enthalten. An dieser Stelle sei der SISTRIX-Optimizer (www.sistrix.de/toolbox/optimizer/) für kleine bis mittlere Websites und das bereits zu Beginn angesprochene Monitoring von Audisto (www.audisto.com/monitoring) für größere Websites erwähnt.

Egal ob Sie ein repräsentatives URL-Set wählen oder die gesamte Domain crawlen - es geht immer darum, negative Auswirkungen auf die SEO-Performance durch Änderungen an der Website frühzeitig mitzubekommen. Dafür benötigen Sie zusätzlich zu Ihrem Onpage-Monitoring ein Alerting, welches Sie auf Veränderungen aktiv hinweist. Sie müssen demnach eine Benachrichtigung erhalten, sobald eine der definierten Prüfungen nicht erfolgreich ist. Am naheliegendsten ist eine Benachrichtigung per E-Mail. Hier bietet sich die Open-Source-Lösung Grafana an, welche nicht nur ein sehr leistungsfähiges Dashboard-Tool ist, sondern auch Warnungen abschicken kann. Hier können Sie die Prüfungen, ob beispielsweise die Startseite einen Statuscode 200 liefert, durchführen. Grafana selbst stellt verschiedene Schnittstellen zur Verfügung. So kann die Nachricht, dass ein Fehler gefunden wurde, als E-Mail verschickt werden. Bei zu vollen Postfächern bietet sich aber auch eine direkte Nachricht in Microsoft Teams an. Abhängig vom eingesetzten Tool ist aber auch eine Anbindung an das eigene Ticketsystem denkbar. Zudem können Sie idealerweise einstellen, welche Personen bei welcher Prüfung eine Benachrichtigung erhalten sollen. Während eine Änderung an der robots.txt bei einem SEO landen sollte, ist die Nachricht über einen nicht mehr gefundenen Google-Tag-Manager-Code wahrscheinlich eher ein Thema für die IT. Das verhindert gleichzeitig, dass zu viele irrelevante Benachrichtigungen aus dem Monitoring-System im Posteingang auflaufen und dadurch wirklich relevante Mitteilungen untergehen.

Verbindung verschiedener Datenquellen

Auch wenn die Speicherung und Weiterverarbeitung der verschiedenen Datensätze bereits für sich allein viele

Vorteile bieten, entfaltet ein zentrales SEO-Data-Warehouse sein volles Potenzial erst bei der Verknüpfung unterschiedlicher Datensätze. Anstatt verschiedene Datenpools etwa in Excel manuell zu verbinden, sind Sie mit einem Data-Warehouse in der Lage, unterschiedliche Daten dynamisch und nach Bedarf zu kombinieren. Grundlage dafür ist ein gemeinsamer Schlüssel, mit dem Sie die verschiedenen Daten auf Basis eines gemeinsamen Merkmals zusammenführen. Im Google DataStudio ist diese Funktion unter den Begriffen „Data Blending“ und „Join Key“ bekannt.

Ein Beispiel verdeutlicht den ungeheuren Vorteil: In Ihrem Data-Warehouse speichern Sie täglich Leistungsdaten aus der Google Search Console, Umsatzdaten aus Google Analytics und Sichtbarkeitsdaten aus SISTRIX für jede URL. Anstatt sich beim nächsten Sichtbarkeitseinbruch mit mühsamen Exporten und SVERWEISEN die Auswirkungen auf Traffic und Umsatz zusammenzubauen, ermöglicht Ihr Data-Warehouse, alle Datenpunkte in einem Bericht in Relation zu setzen. In diesem Fall ist der gemeinsame Schlüssel die URL. So können Sie Fragen beantworten wie:

- » Hatte der Sichtbarkeitsverlauf einen Einfluss auf den Traffic? Oder sind nur Rankings verloren gegangen, die Ihnen ohnehin keine Klicks brachten?
- » In welchen Verzeichnissen sind Sichtbarkeit und Traffic gefallen? Und wo ist nur die Sichtbarkeit gefallen?
- » In welchen Seitentypen sind Sichtbarkeit und Traffic gefallen, aber der Umsatz stabil geblieben? Sind vielleicht nur Rankings verloren gegangen, die ohnehin nicht zu Konversionen geführt haben?
- » Welche Korrelationen zeigen sich zwischen welchen Metriken in welchen Segmenten? Korreliert die Sichtbarkeit mit realen Impressio-

HINWEIS

Das Open-Source DAWIS von mindshape ist zur freien Verfügung unter <http://einfach.st/dawis> downloadbar. Dort finden Sie unter anderem auch eine virtuelle Maschine mit Anleitung zum Herunterladen für ein erstes Ausprobieren. Damit läuft dann auf dem eigenen Rechner ein kleiner Server für DAWIS.

nen?

Besonders nützlich: Durch den gemeinsamen Schlüssel können Sie Segmentierungen, die Sie beispielsweise für Google-Search-Console-Daten eingerichtet haben, auf andere Datensätze wie die Sichtbarkeitsdaten übertragen.

Ebenfalls spannend ist die Verbindung verschiedener Datensätze mit Logfile-Daten. Mithilfe des DAWIS-Zusatztools „Botplorer“ (derzeit noch in der Beta-Phase) lassen sich die Botzugriffe unabhängig von den Serverlogfiles direkt in Google BigQuery abspeichern. Von dort aus können Sie als Teil des Data-Warehouse beliebig weiterverwendet werden. Sie planen einen Relaunch? Mit der Verbindung aus Logfiles, Klicks und Impressionen aus der Google Search Console sowie Seitenaufrufen aus Analytics verfügen Sie über ein ideales Monitoring. Nutzen Sie die Kombination der Daten für Berichte wie:

- » Zugriffe des Googlebots auf 404-Seiten vs. Seitenaufrufe von 404-Seiten (siehe Abbildung 12): Haben Sie an alle wichtigen Weiterleitungen gedacht oder tauchen in der Tabelle URLs auf, die dort nicht hingehören? Hat ein Nutzer oder der Googlebot die URL aufgerufen? Handelt es sich um ein aktuelles Problem für Ihre Nutzer oder eine historische URL, die nur noch dem Googlebot bekannt ist?
- » Zugriffe des Googlebots auf 301-Seiten vs. Impressionen von 301-Seiten: Wie werden die Weiterleitungen verarbeitet? Welche weitergeleiteten

URLs erzeugen noch Impressionen, obwohl sie gecrawlt wurden? Welche weitergeleiteten URLs wurden noch nicht gecrawlt? Was könnte der Grund sein?

- » Zugriffe des Googlebots auf 200-Seiten vs. Impressionen von 200-Seiten: Welche neuen URLs wurden bereits häufiger gecrawlt, aber haben noch keine Impressionen erzeugt? Stimmt alles mit der Indexierung?
- » Zugriffe auf 200-Seiten nach Google-Search-Console-Cluster: Welche Verzeichnisse/Seitentypen werden stärker oder schwächer gecrawlt? Wie entwickelt sich das Crawling in den letzten Tagen?

Natürlich können Sie die Verbindung von Daten nicht nur zum Monitoring nutzen. Auch SEO-Potenziale lassen sich auf diese Weise aufspüren:

- » Welche URLs haben viele Impressionen aus Non-Brand-Suchanfragen, aber wenig Klicks und kaum Googlebot-Zugriffe? Diese Seiten erzeugen bereits Rankings zu populären Keywords, befinden sich aber anscheinend noch nicht in klickrelevanten Bereichen der SERPs. Prüfen Sie, ob Sie diese Seiten mit besserer interner Verlinkung nach oben bringen können – geringes Crawling ist oft ein Zeichen für schlechte interne Verlinkung.
- » Welche URLs haben viele Klicks, aber weisen laut Analytics-Daten kaum Interaktionen bei organischen Zugriffen auf? Passen Snippet und Inhalt der Seite zueinander oder können Sie hier optimieren?
- » Welche URLs werden vom Googlebot gecrawlt, aber erzeugen keinerlei Impressionen? Müssen diese Seiten für den Googlebot erreichbar sein oder können Sie hier Crawl-Budget einsparen?
- » Welche URLs erzeugen viele Klicks, aber haben schlechte Web-Vital-

Werte? Legen Sie besonderen Fokus auf die Performance-Optimierung dieser Seiten, da dort zahlreiche Nutzer einsteigen. Enttäuscht dann die User Experience und führt zum Absprung, lassen Sie hier eigentlich bereits erreichten SEO-Erfolg verpuffen. Dieser fehlt Ihnen später im Reporting an Ihre Kunden oder Vorgesetzten.

Das ist erst der Anfang

Die vorgestellten Anwendungsfälle sind nur ein kleiner Teil dessen, was mit einem umfassenden SEO-Data-Warehouse möglich ist. Eine der größten Stärken liegt in der flexiblen Individualisierbarkeit. Denn durch den Zugriff auf verschiedenste Quellen und Arten

von Datensätzen können Sie diese so verändern und kombinieren, dass sie zu Ihrer projektspezifischen Fragestellung passen. Jedes SEO-Projekt ist individuell und entsprechend flexibel muss auch die Datengrundlage sein. Und jetzt, wo die unterschiedlichsten Daten alle an einem zentralen Ort verfügbar sind, werden auch Ansätze wie Machine Learning und künstliche Intelligenz überhaupt erst möglich. Das zeigt, wo die Reise hingehen wird: Stellen Sie sich vor, Ihre Keywords in der Google Search Console werden automatisch nach Ihrer Suchintention geclustert. Oder Sie können auf die Frage „Wie viel Traffic-Steigerung im SEO ist für kommandes Jahr realistisch?“ endlich eine maschinengestützte, datengetriebene

Antwort geben. Durch die Integration der Daten in die Cloudplattformen der zuvor genannten Anbieter sind Rechenpower und Technologie nicht mehr der beschneidende Faktor. Der erste Schritt ist allerdings immer, die Daten zentral vorrätig zu haben. Das löst ein SEO-Data-Warehouse.

Bestimmt haben auch Sie schon viele Ideen, wie die stärkere Nutzung von Daten im SEO Ihren Arbeitsalltag erleichtern kann. Genau deswegen ist DAWIS als Open-Source-Projekt veröffentlicht. Bringen Sie gerne Ihre Ideen mit ein oder beteiligen Sie sich selbst direkt an der Weiterentwicklung. Für mehr Spaß an der Arbeit, stärkere Daten und besseres und nachhaltigeres SEO! ¶

IMPRESSUM

Herausgeber & Chefredakteur (verantwortlich):

Mario Fischer

E-Mail: redaktion@websiteboosting.com

Autoren dieser Ausgabe:

Dr. Martin Bahr, Alexander Beck,
Stephan Czysch, Darius Erdt,
Sebastin Erhofer, Tobias Häring,
Juliane Höger, Patrick Lürwer,
Annika Maroth, Rico Melzer,
Magdalena Mues, Malte Müller,
André Preuksatz, Jannis Uch,
Patricia Unfried, Alexandra Zytkowskiak

Anzeigenleitung:

Markus Lutz

E-Mail: anzeigenleitung@websiteboosting.com

Art Direction, Layout/Produktion:

Kai Neugebauer

Lektorat:

Ursula Wenke, www.lektorat-wenke.de

Fotos & Illustrationen:

Website Boosting / GettyImages
Titel: ilbusca/gettyimages.de

Druck:

Vogel Druck und Medienservice GmbH
Leibnizstr. 5, 97204 Höchberg

Vertrieb:

PressUp GmbH
Postfach 70 13 11
22013 Hamburg

E-Mail: websiteboosting@pressup.de

Abonnement:

Website Boosting Aboservice
PressUp GmbH
Postfach 70 13 11
22013 Hamburg
Tel. 040 / 38 6666 - 342
Fax: 040 / 38 6666 - 299

E-Mail: websiteboosting@pressup.de

Erscheinungsweise: 6 x jährlich

Bezugspreis: Einzelheft: 11,80€

Bezugspreis Inland jährlich 62,00€ inkl. Versand

Bezugspreis Ausland jährlich 70,80€
inkl. Versand

Studenten im Inland erhalten gegen Vorlage einer
Immatrikulationsbescheinigung einen
Preisvorteil – Details finden Sie auf der Website.

Verlagsleitung:

Michael Müßig

Tel: +49 931 / 26 038 04,

verlag@websiteboosting.com

Anschrift des Verlages

Hotspot Verlag GmbH
Obere Landwehr 4a, 97204 Höchberg

Tel: + 49 931 / 26 038 04

Fax: +49 931 / 26 038 05

E-Mail: verlag@hotspotverlag.de

www.hotspotverlag.de

Geschäftsführung:

Kai Neugebauer

Die Inhaber- und Beteiligungsverhältnisse
lauten wie folgt: Gesellschafter zu 100%
ist die Webvalue Holding GmbH

ISSN: 2191-6241

Für unverlangt eingereichte Texte und Daten kann keine Haftung übernommen werden. Sämtliche Veröffentlichungen in Website Boosting erfolgen ohne Berücksichtigung eines eventuellen Patentschutzes. Markennamen werden ohne Gewährleistung einer freien Verwendung benutzt. Trotz sorgfältiger Recherche kann für die Richtigkeit des Inhalts keine Haftung übernommen werden. Namentlich gekennzeichnete Artikel geben nicht unbedingt die Meinung der Redaktion wieder.