

Stephan Czysch

# Besser suchen & finden mit regulären Ausdrücken:

## RegEx im Google-Search-Console-Leistungsbericht



### DER AUTOR



**Stephans** Leidenschaft gilt besseren Websites. Er beschäftigt sich intensiv mit dem Thema datengetriebenes Online-Marketing und arbeitet aktuell an einem Tool rund um die Google Search Console. Stephan spricht regelmäßig auf Konferenzen zu Themen wie Online-Marketing-Strategien und datengetriebenes SEO. Sein Wissen teilt er zudem in Fachartikeln und Büchern sowie in Seminaren.

Im Leistungsbericht stecken für viele Online-Marketer die spannendsten Daten der Google Search Console. Im Bericht lässt sich – wenn aufgrund ausgeblendeter Suchanfragen auch nicht vollständig – nachvollziehen, über welche Suchanfragen Nutzer auf die eigene Website gekommen sind. Denn in Webanalyse-Tools steht als Suchanfrage fast ausschließlich (not provided).

Für die Optimierung der Website ist das Verständnis der Suchanfragen und Nutzerintention enorm wichtig. Bisher konnten die Daten im Leistungsbericht anhand von „ist exakt“, „enthält“ oder „enthält nicht“ gefiltert werden – seit April 2021 steht zusätzlich die Filterung mittels regulärer Ausdrücke (Englisch: regular expressions, kurz RegEx) zur Verfügung. Somit haben Sie ab sofort ein sehr mächtiges Filterwerkzeug zur Verfügung, mit dem praktisch auch sehr komplexe (Filter-)Abfragen möglich werden. Der Nachteil? „RegEx“ sieht auf den ersten Blick deutlich komplizierter aus, als es tatsächlich ist – wie man beim Selbst-Ausprobieren recht schnell erkennen kann.

Welche Möglichkeiten sich durch die Einführung dieser Neuerung ergeben und wie Sie diese selbst nutzbringend anwenden können, zeigt Ihnen Stephan Czysch.

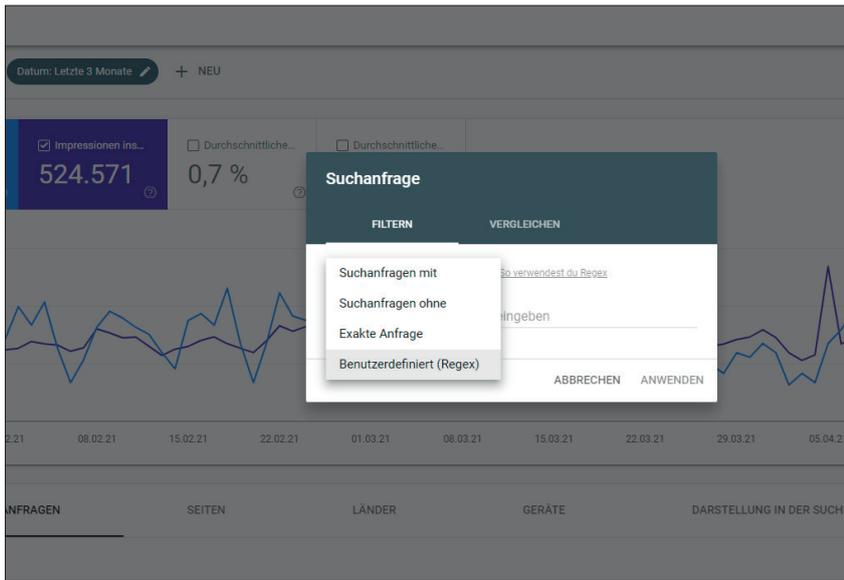


Abb. 1: In der Google Search Console können sowohl Suchanfragen als auch Adressen anhand regulärer Ausdrücke durchgesucht werden; zudem steht die Funktion für Vergleiche zur Verfügung

Im Sommer 2020 gab es in der Google-Search-Console-Hilfe ein kurzzeitiges Mysterium zu entdecken: Dort wurde davon gesprochen, dass der beliebte Leistungsbericht mittels regulärer Ausdrücke durchsucht werden kann. Abseits des Hinweises in der Hilfe gab es in der Search Console selbst jedoch kein Anzeichen dafür, dass diese Möglichkeit besteht. Und so schnell der Hinweis online war, so schnell war er auch wieder verschwunden.

Bis zum 7. April 2021 war es um das Thema RegEx und die Google Search Console ruhig – bis Google im Blogbeitrag „Improved data filtering and comparison on Performance reports“ (<http://einfach.st/gscblog3>) die Funktion öffentlich ankündigte und im Leistungsbericht als neue Filteroption „Benutzerdefiniert (RegEx)“ erschien.

Im Vergleich zu den bisherigen Filtern bieten reguläre Ausdrücke mehr Flexibilität, da z. B. mehrere unterschiedliche Zeichenfolgen gesucht werden können. Das ist unter anderem beim eigenen Markennamen relevant. Ein Beispiel: Nach dem Modehändler „Peek & Cloppenburg“ wird u. a. über Suchanfragen wie „p und c“, „p&c“, „punc“, „peek cloppenburg“ oder auch über Falschschreibweisen wie „peak cloppenburg“ gesucht – all diese Varianten lassen sich

mittels des „Suchanfragen mit“-Filters schwer treffen. Zwar enthalten alle Suchanfragen z. B. ein „p“ am Anfang – doch bei „Suchanfragen mit“ ist es komplett egal, wo das „p“ auftritt. Gezielt nach „Beginnt mit p“ lässt sich im Bericht mit den Standardfiltern also nicht suchen. Und diese Einschränkung würde zudem Suchanfragen wie „pullover“ oder „poloshirt“ einschließen. Unter anderem hierbei helfen reguläre Ausdrücke. Diese zusätzliche Suchoption steht für die Filterung sowohl von Suchanfragen als auch URLs zur Verfügung.

Häufigste Suchanfragen	Impressionen
was ist eine sitemap	132
wie kann man bei google weiter oben erscheinen	79
wie schnell indiziert google	74
wie komme ich bei google auf seite 1	71
was sind gute keywords	65
welche backlinks entwerfen	64
über welche keywords wird meine seite gefunden	64
was ist ein guter sichtbarkeitsindex	63
wie homepage bei google sichtbar machen	63
was sind nofollow backlinks	62

Abb. 2: Mit wenig Aufwand lassen sich über reguläre Ausdrücke unterschiedliche Nutzerfragen ermitteln, zu denen die Website aktuell gefunden wird

## TIPP

Bevor Sie anfangen, die hier gezeigten regulären Ausdrücke abzutippen: Unter <http://einfach.st/scregex> können Sie sich diese komfortabel kopieren. Dort laden Sie sich jeden ein, weitere interessante Ausdrücke als Kommentar zu hinterlassen.

## Ein paar handliche RegEx für Ihre Analysen

Als Einstieg in die Welt der regulären Ausdrücke werfen Sie am besten einen Blick auf meinen Beitrag in der Website Boosting #60. Und auch wenn reguläre Ausdrücke auf den ersten Blick wie ein Buch mit sieben Siegeln erscheinen: Bleiben Sie am Ball und arbeiten sich nach und nach in das Thema ein. Durch Übung und Hilfetexte im Netz lässt sich für (fast) jedes Problem eine Lösung finden. Bei regulären Ausdrücken gibt es in der Regel mehr als eine mögliche Umsetzung. Doch dazu später mehr.

## Suchanfragen mit dem eigenen Markennamen finden

Um verschiedene Schreibweisen des eigenen Markennamens zu finden, kann die „oder“-Verknüpfung verwendet werden. Das „oder“ wird bei regulären Ausdrücken durch „|“ definiert.

Für Peek & Cloppenburg sieht ein (vermutlich unvollständiger) regulärer Ausdruck für die Anzeige von Markensuchen also so aus:

```
pc|peek|peak|cloppenburg-|pundo
```

Dieselbe Logik kann zudem verwendet werden, um z. B. nach allen Suchanfragen zu filtern, die „kleider“ oder „hosen“ beinhalten.

```
kleider|hosen
```

### Suchfragen mit Fragen finden

Wer die Clusterung von Suchanfragen weiterdenkt, der landet schnell bei Fragewörtern. Auch diese lassen sich über einen RegEx mit wenig Aufwand identifizieren. Einfach die unterschiedlichen Fragewörter miteinander kombinieren und schon liefert Google im Leistungsbericht die Daten zurück.

Schauen wir in diesem Zusammenhang einmal auf das Thema, dass unterschiedliche reguläre Ausdrücke dasselbe Ergebnis liefern. So liefert z. B.

```
wer .*|wem .*|wen .*|wessen
.*|wie .*|wann .*|wo .*|welche
.*|was .*|wobei .*|womit
.*|woran .*|wohin .*|wobei
.*|weshalb .*|warum .*|wieso
.*|worauf .*|worum .*|wovor
.*|wodurch .*|woher .*|weswegen
.*|woraus .*
```

eine ganze Reihe an Suchanfragen mit Fragewörtern. Das Leerzeichen nach jedem Fragewort benötigen wir, da eine Suche nach „wer“ auch „Werbung“ oder „Wertung“ zurückliefern könnte. Das „.\*“ sorgt dafür, dass jedes beliebige Zeichen (.) in beliebiger Häufigkeit (\*) nach dem Leerzeichen kommen kann.

Doch ist dieser reguläre Ausdruck bereits fehlerfrei? Aktuell wird nicht definiert, ob vor z. B. dem „wer“ noch

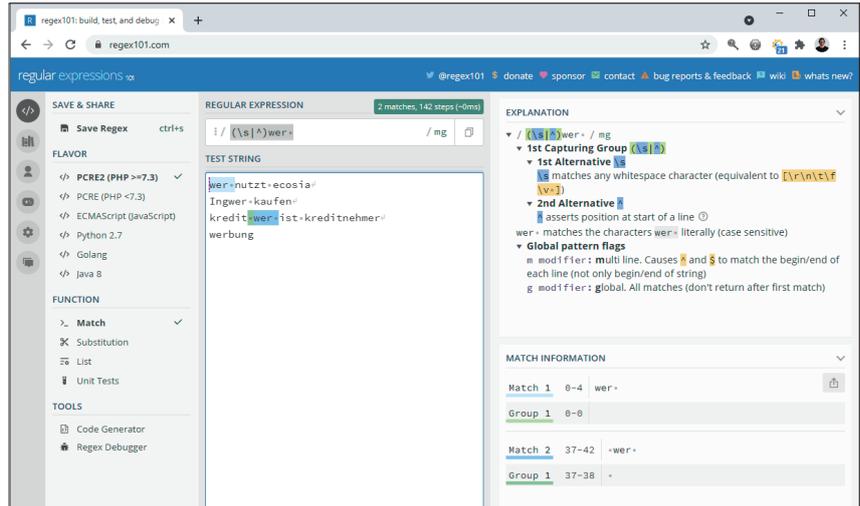


Abb. 3: Da vor der Zeichenfolge „wer“ nur ein Leerzeichen oder nichts kommen darf, ist „Ingwer kaufen“ kein Treffer; auch „werbung“ wird nicht erfasst, da der reguläre Ausdruck zwingend ein Leerzeichen nach dem Fragewort erfordert

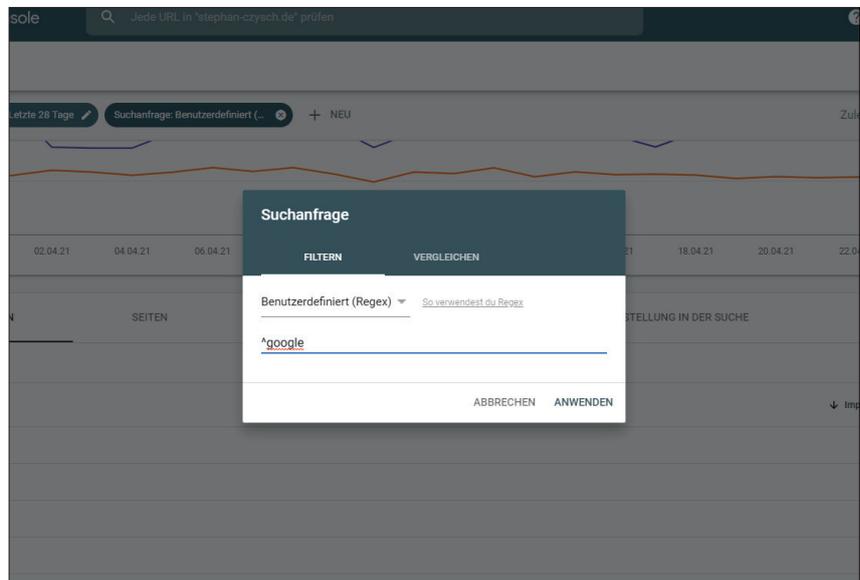


Abb. 4: Durch die Angabe des ^ vor der gesuchten Zeichenfolge wird definiert, dass die anzuzeigenden Suchanfragen mit der gewünschten Zeichenfolge google beginnen müssen

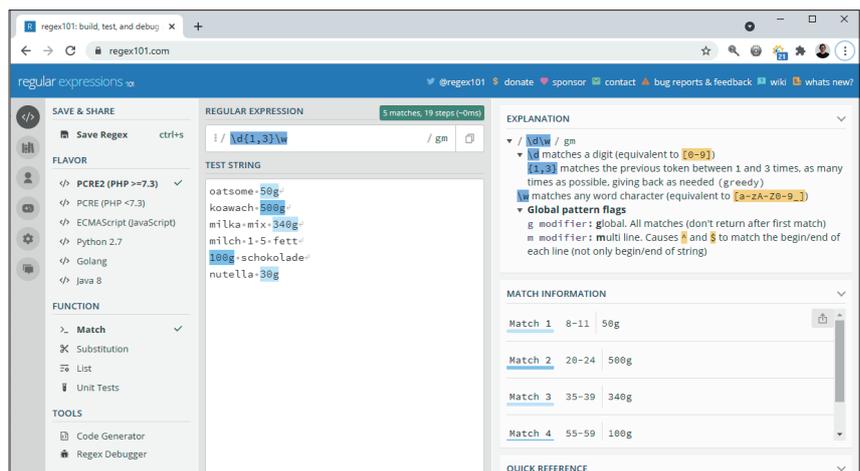


Abb. 5: In diesem Beispiel wurde nach 1 bis 3 aufeinanderfolgenden Ziffern gesucht – und zusätzlich definiert, dass direkt hinter der Ziffer ein Buchstabe kommen muss

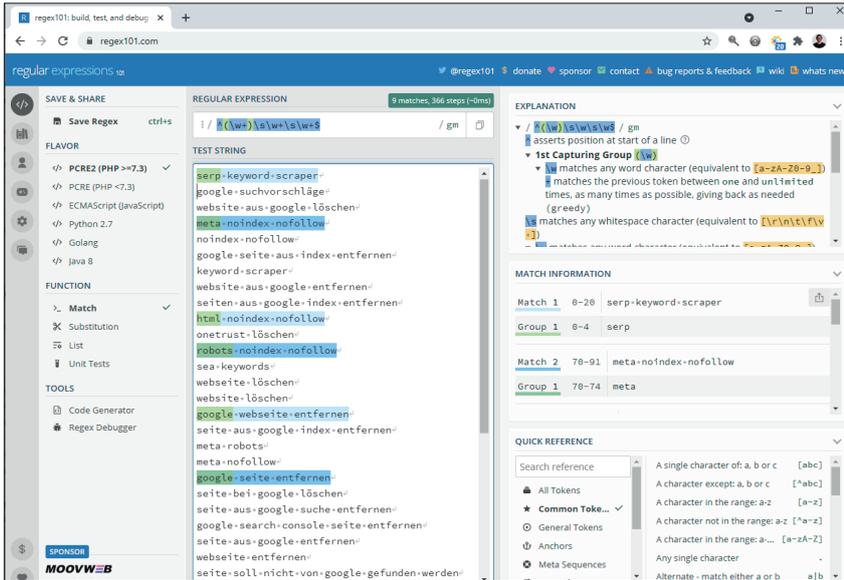


Abb. 6: Durch den Ausdruck `^((w+)|s)w+|s|w+$` werden Drei-Wort-Suchanfragen als Ergebnis geliefert

etwas kommen darf oder nicht. Und ein „wer“ steckt z. B. in „Ingwer kaufen“ drin – das ist allerdings keine Frage.

Wie kann der reguläre Ausdruck also verbessert werden? Wir möchten Suchanfragen angezeigt bekommen, die entweder mit „wer“ beginnen oder bei denen vor dem „wer“ maximal ein Leerzeichen kommt. Es ist nämlich nicht immer der Fall, dass das Fragewort am Anfang steht. Die Suchanfrage „über welche keywords wird meine seite gefunden“ ist dafür ein Beispiel. Um zu definieren, dass vor dem Fragewort entweder nichts oder ein Leerzeichen steht, können wir auf die Angabe „(\s|^)“ zurückgreifen. Das \s steht dabei für Leerzeichen, und das ^ definiert, dass vor dem „wer“ alternativ nichts kommen darf. Diese beiden Angaben mit dem bekannten oder („|“) verknüpfen und schon wurde der bisherige Ausdruck verbessert.

```
(\s|^)wer .*|wem .*|wen
.*|wessen .*|wie .*|wann .*|wo
.*|welche .*|was .*|wobei
.*|womit .*|woran .*|wohin
.*|wobei .*|weshalb .*|warum
.*|wieso .*|worauf .*|worum
.*|wovor .*|wodurch .*|woher
.*|weswegen .*|woraus .*
```

Wenn wir uns nochmals den bisherigen regulären Ausdruck vor Augen führen, dann fällt auf, dass alle Zeichenketten mit einem w beginnen. Kein Wunder, wir suchen schließlich nach W-Fragen. Anstatt „wer“ oder „was“ zu suchen, könnten wir das „w“ auch einmalig als zwingend definieren. Das machen wir, indem wir das „w“ ganz nach vorne stellen und direkt eine runde Klammer anschließen. Zusammengesetzt ergibt sich aus den einzelnen Angaben folglich „wer“, „wem“, „wessen“ etc. Die nächste Verbesserung ist:

```
(\s|^)w(er .*|em .*|en
.*|essen .*|ie .*|ann .*|o
.*|elche .*|as .*|obei .*|omit
.*|oran .*|ohin .*|obei
.*|eshalb .*|arum .*|ieso
.*|orauf .*|orum .*|ovor
.*|odurch .*|oher .*|eswegen
.*|oraus .*)
```

Doch auch dieser Ausdruck kann noch verschönert werden, denn hinter jedem unserer nun zusammengesetzten Fragewörter haben wir ein Leerzeichen als zwingend erforderlich angegeben. Dem Leerzeichen können sich durch die Angabe „\*“ beliebige andere Zeichen mit einer Häufigkeit von 0 bis unendlich anschließen. Um nach einem Leerzeichen zu suchen, können wir alternativ auf den bereits bekannten Operator \s zurückgreifen. Und das „\*“ brauchen wir für unseren Zweck nicht zwingend. Das Ergebnis ist:

```
(\s|^)w(er|em|en|essen|ie|an-
n|o|elche|as|obei|omit|oran|o-
hin|obei|eshalb|arum|ieso|o-
rauf|orum|ovor|odurch|oher-
|eswegen|oraus)\s
```

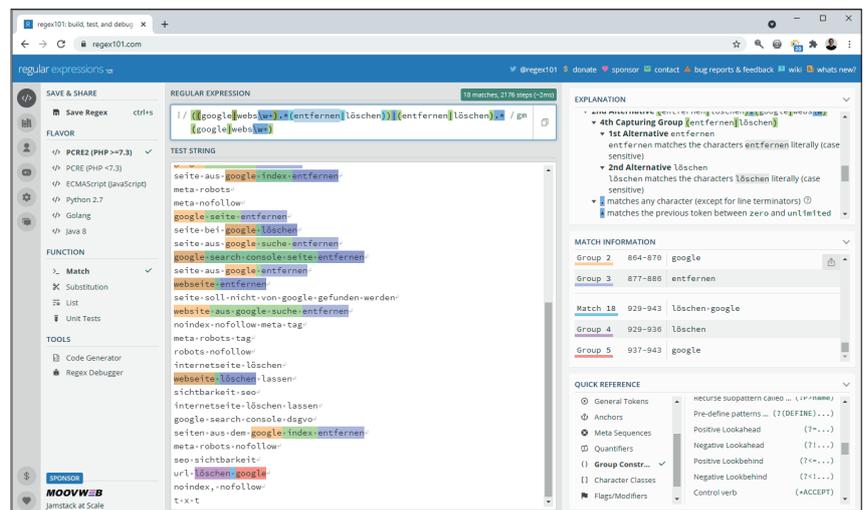


Abb. 7: Hier wird es schon etwas komplexer: Google oder Website sollen zusammen mit entfernen oder löschen in den Suchanfragen vorkommen – unabhängig davon, welche der beiden Wortgruppen zuerst vorkommt

### Transaktionale Suchanfragen identifizieren

Die Verknüpfung mit „oder“ kann allerdings auch für ganz andere Zwecke eingesetzt werden. Wen es z. B. interessiert, welche Suchanfragen „transaktional“ sind, also Wörter wie kaufen oder bestellen beinhalten, der greift ebenfalls zu dieser Verknüpfung. Eine Suche nach

`kauf|bestellen`

zeigt beispielsweise die entsprechenden Suchanfragen an.

### Suchanfragen finden, die mit einer bestimmten Zeichenfolge beginnen

Einer der großen Vorteile der RegEx-Suchfunktion ist, dass definiert werden kann, wo eine Zeichenfolge vorkommen soll. Dazu wird das `^` der Zeichenfolge vorangestellt. Durch die Filterung nach

`^google`

werden nur die Suchbegriffe zurückgeliefert, die mit google beginnen. Da nicht definiert wird, ob und falls ja, was sich dieser Zeichenfolge anschließen soll, ist auch „googlebot“ ein möglicher Treffer.

### Suchanfragen finden, die auf eine bestimmte Zeichenfolge enden

Wenn die gewünschte Zeichenfolge am Ende der Suchanfrage stehen soll, dann ist die Nutzung von `$` der richtige Operator. Durch den Ausdruck

`google$`

erscheinen nur Suchanfragen, bei denen google am Ende der Anfrage steht.

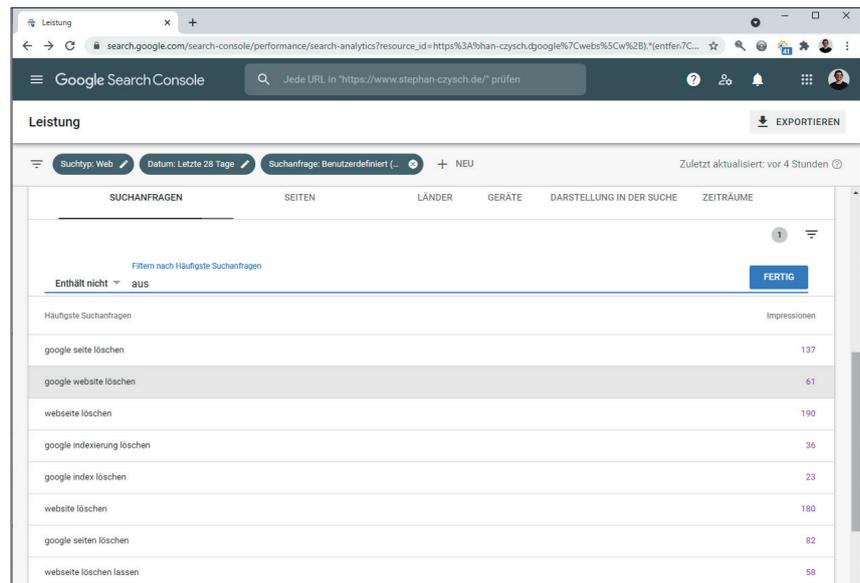


Abb. 8: Durch den Trichter oberhalb der Datentabelle lassen sich die angezeigten Daten nochmals filtern – im Beispiel werden alle Suchanfragen, die „aus“ beinhalten, ausgeblendet

### Suchanfragen finden, die aufeinanderfolgende Zahlen enthalten

Es kommt immer wieder vor, dass Nutzer in einer Suchanfrage z. B. das aktuelle Jahr oder Abmessungen bzw. Größen angeben. Während sich ein einzelnes Jahr leicht mit den bisherigen Filtern finden lässt, ist die Suche anhand von Ziffern flexibler – wahlweise mit einer gewünschten Häufigkeit. Für die Suche nach Ziffern kann „`d`“ verwendet werden, für die gesuchte Häufigkeit die Angabe `{min,max}`. Die Angabe

`\d{1,2}`

führt z. B. dazu, dass eine ein- oder zweistellige Zahl Teil der Treffermenge ist. Auch hier gibt es wie gewohnt unterschiedliche Möglichkeiten, dies in einem regulären Ausdruck zu beschreiben. So kann statt `\d` auch der Bereich von `[0-9]` angegeben werden, damit eine Zahl zwischen 0 bis 9 gesucht wird.

Über reguläre Ausdrücke lässt sich mit geringem Aufwand z. B. nach Suchanfragen suchen, die Gewichtsangaben beinhalten. Denken Sie hier an Keywords rund um Lebensmittel. Im

Beispiel aus Abbildung 5 wurde nach

`\d{1,3}\w`

gesucht, um Suchbegriffe anzuzeigen, die aus 1 bis 3 Ziffern bestehen und denen sich ein Buchstabe anschließt. Entsprechend fällt „milch 1 5 fett“ durch das Suchraster – denn an die Ziffern schließt sich nicht direkt ein Buchstabe, sondern ein Leerzeichen an.

### Suchanfragen mit einer bestimmten Wortanzahl finden

Dank regulärer Ausdrücke ist es problemlos möglich, Suchanfragen mit einer bestimmten Wortanzahl zu finden. Hierzu wird das Leerzeichen als Indikator genommen. So findet

`^(\\w+)\s\\w+\s\\w+$`

alle Suchanfragen mit drei Wörtern. Die Angabe `\w` steht dabei für einen beliebigen Buchstaben oder eine beliebige Zahl und durch das `+` wird definiert, dass mindestens ein Buchstabe oder eine Zahl vorkommen muss. Das Dollarzeichen am Ende definiert, dass hinter dem letzten „Wort“ nichts Weiteres kommen darf. Ohne diese Angabe

würden auch Suchanfragen mit mehr als drei Wörtern angezeigt.

### Entweder Zeichenfolge A oder B und Zeichenfolge C oder D – oder in umgekehrter Reihenfolge

In der Beispiel-Keywordliste kommen die Themen „löschen“ bzw. „entfernen“ sowie „Google“ und „Website“ immer wieder in unterschiedlichen Kombinationen zusammen vor. Auch diese Begriffe lassen sich mit etwas Kombinatorik finden:

```
((google|webs\w+).*(entfernen|löschen))|(entfernen|löschen).*(google|webs\w+)
```

Dieser reguläre Ausdruck besagt, dass die Zeichenfolge „google“ oder „webs“ (um Website oder Webseite zu matchen) enthalten sein muss und dass sich diesen Begriffen entweder „entfernen“ oder „löschen“ anschließen soll. Zudem wurde die Syntax noch mal umgedreht, da auch „url löschen google“ in die Trefferliste mit aufgenommen werden soll.

An diese Stelle möchte ich auf den Tabellenfilter oberhalb der Datentabelle hinweisen. Durch diesen können irrelevante Zeilen einfach ausblendet werden. Zur Erinnerung: Der Filter oberhalb des Charts des Leistungsberichts definiert, welche Daten abgefragt werden sollen. Diese erscheinen anschließend in der Datentabelle unterhalb des Charts – und für die Tabelle gibt es einen separaten Filter. Dieser Filter verändert nicht die Anfrage, sondern blendet Zeilen ein oder aus.

### Leben wir nun in der perfekten Google-Search-Console-Welt?

Manchmal treibt – zumindest mich – die RegEx-Filterung der Suchanalysedaten zur Weißglut. Nicht nur,

dass ich bei meinen Ausdrücken etwas übersehe oder vergesse: Es werden vor allem nicht alle Ausdrücke unterstützt, da Google die RE2-Syntax nutzt. Welche Angaben aktuell unterstützt werden, ist unter <http://einfach.st/github43> nachlesbar. Am schmerzlichsten wird eine Möglichkeit des Ausschlusses vermisst, also z. B. alle Suchanfragen anzuzeigen, die nicht den eigenen Markennamen betreffen. Ich habe viel rumprobiert, aber keine Lösung dafür gefunden. Wer eine hat: immer her damit! :)

Während in Google Analytics über die Auswahl von „einschließen“ oder „ausschließen“ eine einfache Möglichkeit vorhanden ist, genau das Gegenteil eines aktuellen regulären Ausdrucks als Ergebnis zu erhalten, fehlt dies in der Search Console (noch?). Doch wenn alles klappt, dann ist die neue Filterung auf jeden Fall eine Arbeitserleichterung, um sowohl Suchanfragen als eben auch in URLs zu suchen. Da es in den Beispielen zu kurz gekommen ist: Reguläre Ausdrücke können in der Search Console auch innerhalb von Vergleichen für Suchanfragen oder URLs genutzt werden. In diesem Fall werden die Suchanfragen der einzelnen Gruppen allerdings nicht angezeigt. Das macht auch Sinn, da Google bei Vergleichen in der Tabelle die einzelnen Werte vergleichen würde.

Wie gehe ich mit dem Fehlen der „Nicht“-Option um? Ehrlicherweise arbeite ich selten mit dem Leistungsbericht im Interface der Google Search Console und frage die Daten stattdessen über die API ab. Denn das Interface hat seine Schwächen: Neben der Beschränkung auf maximal 1.000 Zeilen blicke ich vor allem bei Erstanalysen auf die Gegenüberstellung von Suchanfrage und Einstiegsseite – das geht im Interface nicht.

Und Analysen, die ich regelmäßig in identischer Form wiederhole (Welche Suchanfragen sind Fragen? Wie viele Branded-Keywords gibt es? ...), kann ich außerhalb der GSC einfach besser über Vorlagen durchführen. Jedes Mal die gewünschte Konfiguration in der Search Console wieder zusammenklicken? Nein, dann lieber direkt die Leistungsdaten über die API abfragen, anreichern und dann weiterarbeiten. Über Bookmarklets lassen sich zwar direkte Einstiege in der Google Search Console bauen und gewisse Einstellungen direkt hinterlegen, aber irgendwie ist es das nicht für mich.

Infokasten: Über Bookmarklets lassen sich Arbeitsschritte abkürzen. So ist es z. B. möglich, den Domainnamen der gerade aufgerufenen Webadresse auszulesen und an die Google Search Console oder viele beliebige SEO-Tools zu übergeben. Ein Klick, und schon erscheinen die z. B. auf Deutschland vorgefilterten Suchanalyse-Daten der Domain oder einzelnen URL auf dem Bildschirm. Unter <http://einfach.st/stefanczysch> sind ein paar der Bookmarklets zu finden, die mir die Arbeit erleichtern.

Das Schöne: Es gibt eine ganze Reihe an fertigen Guides, um mit wenig Aufwand die Google-Search-Console-API anzuzapfen. Und bei Guides hört es nicht auf: Mit kostenfreien Tools wie Search Analytics for Sheets (<https://searchanalyticsforsheets.com/>) lassen sich die Suchanalyse-Daten für die eigene Website in Windeseile und ohne jegliche Programmierkenntnisse in Google Sheets importieren. Liegen die Daten erst mal in Tabellenform vor, dann lassen sie sich einfach anreichern. Google Sheets kann hier großartig helfen, da das kostenlose Tabellenprogramm reguläre Ausdrücke unterstützt!¶