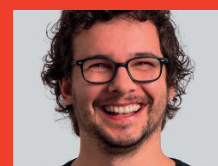


Stephan Czysch

RegEx – besser suchen & finden. So nutzen Sie „reguläre Ausdrücke“ richtig!

Sind Sie erst kürzlich auf einer Konferenz oder beim Lesen eines Artikels über den Begriff „RegEx“ gestolpert und haben sich gefragt, was das eigentlich ist? In diesem Beitrag erfahren Sie praxisnah und einsteigerfreundlich, wie reguläre Ausdrücke (englisch: regular expressions, kurz: RegEx) Ihren Arbeitsalltag vereinfachen können. Denn ob Webanalyse, SEO-Tools oder Ersetzungen in Texten: Mit dem passenden regulären Ausdruck kommen Sie schnell(er) ans Ziel. Experte Stephan Czysch erklärt Ihnen, wie.

DER AUTOR



Stephan Czysch ist als Autor zweier SEO-Fachbücher und als Speaker auf Konferenzen bekannt. Die von Stephan gegründete Online-Marketing-Agentur Trust Agents ist mittlerweile in Dept (deptagency.com) aufgegangen. Er leitet als Geschäftsführer das Digital-Marketing-Team von Dept in Berlin.

Foto: anyaberkut / gettyimages.de

Es gibt immer wieder Dinge, die man gerne schon früher gekannt und eingesetzt hätte. Bei mir war dies unter anderem der Fall, als ich die vielfältigen Einsatzmöglichkeiten von regulären Ausdrücken für meine tägliche Arbeit durchdrungen hatte. Denn wenn es darum geht, bestimmte Zeichenketten zu finden, dann sind reguläre Ausdrücke solchen Alternativen wie Filterungen oder Suchen in Excel-Tabellen oder

Word-Dateien überlegen. Zudem lassen sich RegEx wunderbar dafür einsetzen, um beispielsweise Nutzereingaben wie E-Mail-Adressen in Website-Formularen zu validieren, Ersetzungen vorzunehmen oder um gefundenen Text zu extrahieren. Das klingt nach einer spannenden Fähigkeit, die auch Ihre Arbeit erleichtern kann? Dann lassen Sie uns die Möglichkeiten von regulären Ausdrücken erkunden!

fx	Linkquelle	Domain	Domain enthält SEO oder Mario	.de zu .com ersetzen
1	Linkquelle	Domain	Domain enthält SEO oder Mario	.de zu .com ersetzen
2	https://fischerlaender.de/	fischerlaender.de	FALSCH	https://fischerlaender.com/
3	https://www.website-boosting.de/	website-boosting.de	FALSCH	https://www.website-boosting.com/
4	https://www.seo-suedwest.de/	seo-suedwest.de	WAHR	https://www.seo-suedwest.com/
5	http://www.mario-fischer.de/mario-fischer.html	mario-fischer.de	WAHR	http://www.mario-fischer.com/mario-fischer.html
6	https://www.tagseoblog.de/bilder-seo-ebook	tagseoblog.de	WAHR	https://www.tagseoblog.com/bilder-seo-ebook
7	https://www.selbstaendig-im-netz.de/meine-online	selbstaendig-im-netz.de	FALSCH	https://www.selbstaendig-im-netz.com/meine-online
8	https://www.deptagency.com/de-de/	deptagency.com	FALSCH	https://www.deptagency.com/de-de/
9	https://www.trustagents.de/unternehmen/publikati	trustagents.de	FALSCH	https://www.trustagents.com/unternehmen/publikati
10	http://www.timoaden.de/	timoaden.de	FALSCH	http://www.timoaden.com/
11	regulärer Ausdruck:	REGEXEXTRACT(A2;"^(?:https?:\V/)?(?:[^\n]+@)?(?:www\.)?([^\n]+)")	REGEXMATCH(B2;"seo mario")	REGEXREPLACE(A2;"\de/";".com/")
12				
13				
14				

Abb.1: Google Sheets unterstützt sowohl Suchen (RegExMatch), Ersetzen (RegExReplace) als auch Extrahieren (RegExExtract) mittels regulärer Ausdrücke

Was ist unter RegEx zu verstehen, welche Tools unterstützen reguläre Ausdrücke und wofür lassen sie sich einsetzen?

Egal ob von regulären Ausdrücken, der englischen Bezeichnung regular expressions oder deren Kurzformen RegEx und RegExp gesprochen wird: Mit einem RegEx kann innerhalb von Texten sehr flexibel nach Zeichenketten gesucht werden.

Das klingt erst mal wenig spektakulär, da dies Grundfunktionalität von Programmen wie Word und Excel oder Texteditoren ist. Doch RegEx kann deutlich mehr, da die Suche mit Platzhaltern und Modifikatoren präzisiert werden kann. Sie möchten beispielsweise herausfinden, welche Seitentitel Ihrer Website mit einer Zahl beginnen und zudem die Worte „Tipps“, „Tutorial“ oder „How To“ enthalten? Oder aus Webadressen die Domain extrahieren? Mit dem entsprechenden regulären Ausdruck können Sie das Gewünschte schnell finden.

Sie fragen sich jetzt vermutlich, wie ein RegEx aussieht und in welchen Programmen die Suche nach einem solchen Ausdruck möglich ist. Beginnen wir mit der Klärung der zweiten Frage. Die Liste

der Tools ist lang, deshalb stellt die nachfolgende Auflistung nur eine Auswahl dar:

- » Webanalyse-Tools wie Google Analytics
- » Tabellenkalkulationsprogramme wie Google Spreadsheets
- » SEO-Tools wie Sistrix
- » Website-Crawler wie Screaming Frog
- » Texteditoren wie Notepad++

Microsoft Excel unterstützt über die Standardfunktionen, beispielsweise =FINDEN() und =ERSETZEN() und die eingebaute Suchfunktion reguläre Ausdrücke nur bedingt. Die von Excel angebotenen Platzhalter (auch Wildcards genannt) wie der Stern (*) sind zwar mächtig, jedoch im Vergleich nicht so präzise wie die Suchoperatoren von RegEx. Durch das Add-on „SEO Tools for Excel“ (<https://seotoolsforexcel.com/download-seotools/>) lassen sich RegEx-Funktionalitäten in Excel integrieren. Leider gehört diese Funktion zum kostenpflichtigen Teil des Add-on, allerdings wird eine kostenfreie Testversion angeboten. Schauen Sie sich auch gerne Alternativen wie <https://www.domwoodman.com/posts/how-to-get-excel-in-regex/> an.

Da Google Spreadsheets reguläre Ausdrücke standardmäßig unterstützt und zudem kostenfrei angeboten wird,

fiel die Wahl für Abbildung 1 auf dieses Tool. Die drei verschiedenen regulären Ausdrücke sorgen für Folgendes:

- » Domainnamen aus der Linkquelle extrahieren: REGEXEXTRACT(A2;"^(?:https?:\V/)?(?:[^\n]+@)?(?:www\.)?([^\n]+)")
- » Im extrahierten Domainnamen nach seo oder mario suchen: REGEXMATCH(B2;"seo|mario")
- » .de durch .com in der Linkquelle ersetzen: REGEXREPLACE(A2;"\de/";".com/")

Das müssen Sie über reguläre Ausdrücke wissen

Bei der Suche mit regulären Ausdrücken ist es wichtig zu wissen, dass zwischen Groß- und Kleinschreibung unterschieden wird. Dies nennt man „case-sensitivity“. Entsprechend führt die Suche nach einem großen A dazu, dass ein kleines A keinen Treffer darstellt. Dieses Verhalten lässt sich allerdings ändern: In dem im weiteren Verlauf immer wieder gezeigten Tool <https://regexr.com/> wird beispielsweise durch die Angabe von `\i` definiert, dass sowohl die Klein- als auch die Großschreibweise gefunden werden soll. Dies ist in vielen Tools und Programmiersprachen ebenso.

Das Fragezeichen: ?

Durch die Angabe des Fragezeichens wird definiert, dass das vorangegangene Zeichen 0x oder 1x in der gesuchten Zeichenfolge vorkommen soll. Die Suche nach `https?` führt folglich dazu, dass sowohl `http` als auch `https` ein valides Ergebnis sind, da das `s` als optional markiert wurde.

Das Sternchen: *

Wenn das vorangegangene Zeichen 0x oder beliebig häufig vorkommen soll, dann ist das Sternchen die richtige Wahl. Entsprechend für die Suche nach `hall*` dazu, dass dadurch auch „hal“ ein Treffer ist.

Das Plus: +

Wird ein `+` in den Ausdruck integriert, dann muss das vorangegangene Zeichen mindestens 1x vorkommen.

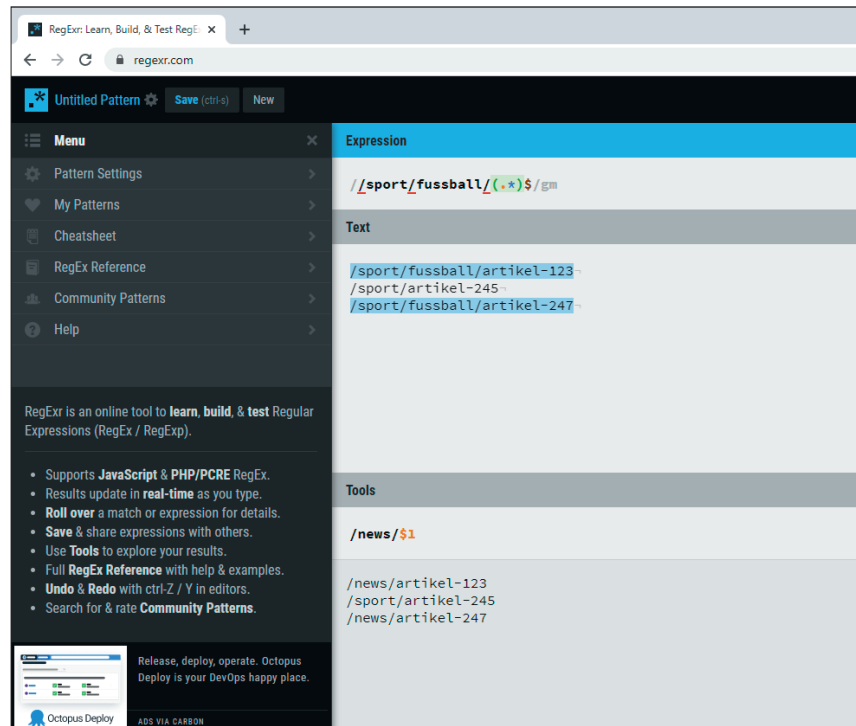


Abb.2: Durch diesen RegEx wird alles nach `/sport/fussball/` Kommende in einer Gruppe zusammengefasst und durch eine Ersetzung (englisch Replace) an `/news/` drangehängt. Über ähnliche Regeln lassen sich z. B. Weiterleitungen in der `.htaccess` umsetzen

NEU

Timme Cloud 2.0

Leistung satt!


TimmeHosting
nginx-Webhosting

Regeln Sie Ihre Cloud-Performance:

- + Jederzeit
- + Zuverlässig
- + Flexibel
- + Skalierbar
- + Stundengenau abgerechnet

timmehosting.de/cloud

SSD

100%

HOSTING
MADE IN
GERMANY

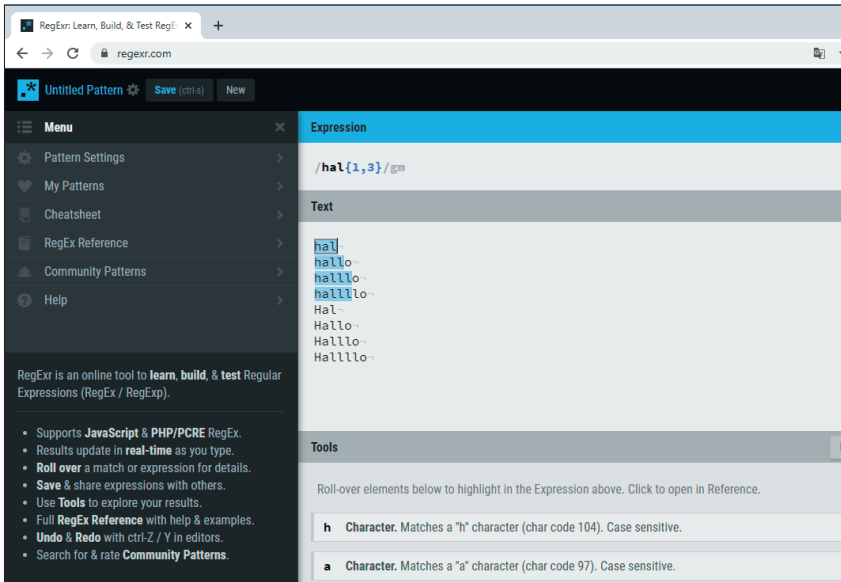


Abb.3: Durch die Angabe von {1,3} wurde definiert, dass das kleine L zwischen ein- und dreimal vorkommen soll. Da nach einem kleinen H gesucht wird, sind die unteren vier Varianten keine Treffer

Entweder a oder b: |

Soll im Suchtext entweder seo oder mario enthalten sein? Dann kann danach über die Verknüpfung der beiden Zeichenfolgen mittels „|“ gesucht werden. Für das konkrete Beispiel bedeutet dies seo|mario.

Beginn eines Suchmusters markieren oder Ausschluss aus der Trefferliste: ^

Um zu markieren, dass die gesuchte Zeichenfolge am Beginn stehen soll, muss dafür das Konsolenzeichen (^) verwendet werden. Durch die Angabe von ^https würde die Zeichenfolge https nur dann als Suchtreffer infrage kommen, wenn diese damit beginnt.

Das Konsolenzeichen kann zudem verwendet werden, um Zeichen aus der Suche auszuschließen. So führt [^abc] dazu, dass diese drei Buchstaben nicht im Suchtreffer vorkommen dürfen. Zur Bedeutung der eckigen Klammern erfahren Sie gleich mehr.

Ende eines Suchmusters markieren: \$

Durch das Dollarzeichen wird angegeben, dass das Suchmuster am Ende der Zeichenfolge vorkommen muss. Suchen Sie z. B. Dateien mit der

Endung jpg, dann können Sie dies über jpg\$ definieren.

Beliebig viele Zeichen: der Punkt .

Der Punkt hat standardmäßig die Bedeutung eines Platzhalters und steht für beliebig viele Zeichen. Wenn die RegEx-Suche nur nach „.“ durchgeführt wird, dann wird der gesamte Text diesem Suchmuster entsprechen.

**Die Standardaktion aufheben: Der Backslash **

Doch was tun, wenn der „.“ nicht für beliebig viele Zeichen stehen soll, sondern z. B. .html gesucht werden soll? Auch das ist natürlich möglich. In diesem Fall muss die normale Bedeutung des Punkts aufgehoben werden – dies geschieht über das Voranstellen eines Backslashes. Um nach .html zu suchen, ist \.html der dazugehörige RegEx.

Gruppen bilden oder exakte Reihenfolge der Zeichen: Die runde Klammer ()

Durch das Umschließen einer Zeichenfolge in runden Klammern muss die dort enthaltene Zeichenfolge in der definierten Reihenfolge vorkommen, damit ein Treffer vorliegt. Dieses Verhalten gilt allerdings auch ohne die Klammern. Wozu also die Klammer nutzen? Um Gruppen zu bilden.

Gruppen sind in regulären Ausdrücken besonders interessant, da diese weiterverarbeitet werden können. Sie können folglich die in runden Klammern beschriebene Zeichenfolge dazu nutzen, um sie mit etwas Neuem zu kombinieren.

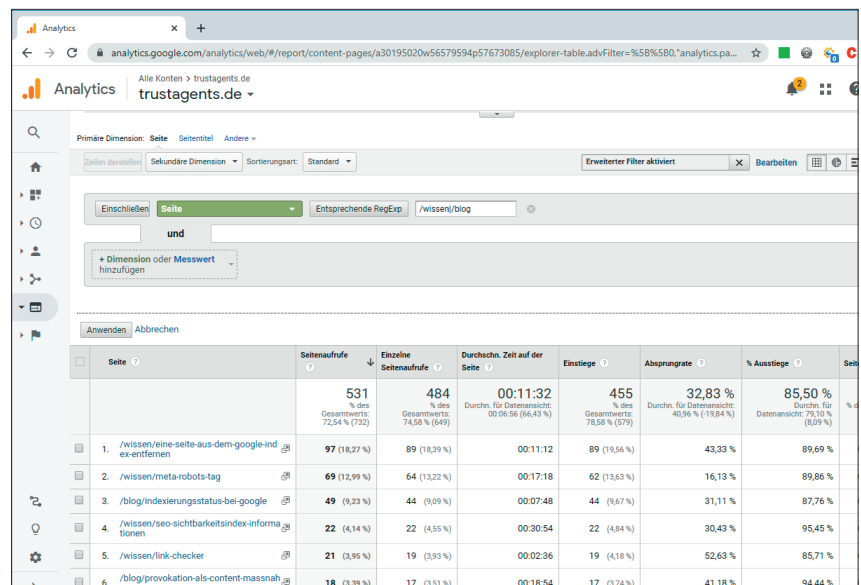


Abb.4: In Google Analytics können die Daten mittels regulärer Ausdrücke nach einem Klick auf „Erweitert“ rechts des Suchfelds eingeschränkt werden. Das Suchfeld ist in der Abbildung durch den gesetzten Filter bereits aktiviert worden

Nehmen wir an, dass sich im Zuge eines Website-Relaunchs die Verzeichnis-Struktur verändert und die Inhalte aus dem Verzeichnis /sport/fussball/ nun in /news/ verschoben wurden. Aus /sport/fussball/artikel-123 wurde folglich /news/artikel-123. Durch den RegEx `/sport/fussball/(.*)$` wird alles nach /sport/fussball/ Kommende als eine Gruppe definiert. Auf diese Gruppen kann in einem Ersetzungsvorgang (englisch: Replace) zugegriffen werden. Die einzelnen Gruppen werden dabei hochgezählt – über \$1 wird die erste über die runden Klammern definierte Gruppe angesprochen, mit \$2 eine ggf. weitere definierte Gruppe und so weiter.

Viele Websites laufen auf der Webserver-Technologie von Apache und verfügen über die .htaccess-Datei. Diese

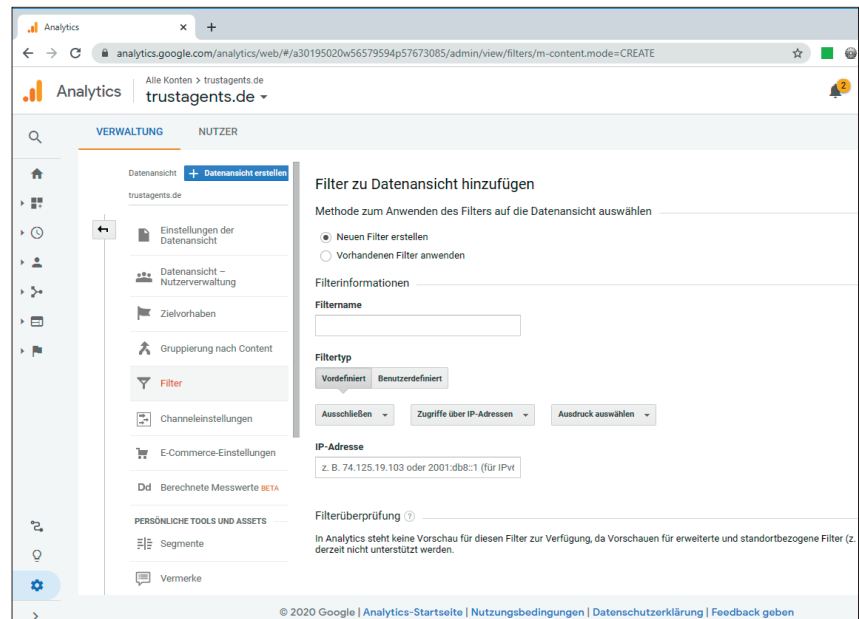


Abb.5: Um z. B. Zugriffe über die eigene (feste) IP-Adresse vollständig aus den Daten auszuschließen, kann in Google Analytics ein Ausschluss-Filter für einzelne Datenansichten definiert werden. Ein regulärer Ausdruck kann bei mehreren verwendeten IP-Adressen die Filterung erleichtern

NEU

Timme Cloud 2.0

Leistung satt!


TimmeHosting
nginx-Webhosting

Cloud-Hosting bietet Ihnen enorm viele Vorteile. Besonders, wenn die Bedienung so einfach ist wie bei unserer neuen Managed Cloud. Sie können ganz intuitiv über Regler die Cloud-Performance einstellen. Die Kosten werden stundengenau – ohne schwankende Kosten für den Traffic - abgerechnet.

Sie erwarten Lastspitzen während einer Kampagne oder eines TV-Auftritts? Mit der Timme Cloud sind Sie bestens gewappnet! Sprechen Sie uns an, wir begleiten Sie schon während der Planungsphase!

timmehosting.de/cloud



kleine und mächtige Datei kann z. B. für das Setzen von Weiterleitungsregeln mit regulären Ausdrücken verwendet werden. Die Datei finden Sie im Hauptverzeichnis Ihrer Website. Um die Beispiel-Adressen von /sport/fussball/ auf /news/ umzuschreiben, muss folgende Zeile in die .htaccess geschrieben werden:

```
RedirectMatch ^/sport/fussball/(.*)$ /news/$1
```

Durch das vorangestellte Konsolenzeichen (^) wird definiert, dass die Adresse mit dem Muster /sport/fussball/ beginnen soll – andernfalls würde die Regel auch für Adressen wie /beliebiges-Verzeichnis/sport/fussball/ gelten.

Bereiche definieren oder beliebige Reihenfolge der Zeichen: Die eckige Klammer []

Wenn das Suchmuster durch eine eckige Klammer umfasst wird, dann müssen diese nicht in der angegebenen Reihenfolge im Text vorkommen. Die Angabe h[ae]llo bedeutet, dass auf ein h entweder ein a oder ein e folgen soll und sich ein „llo“ anschließen muss. Entsprechend sind hallo oder hello Suchtreffer – haello allerdings nicht, da nur das A oder das E vorkommen soll (aber nicht beide).

Die eckige Klammer hat zudem eine weitere Funktion: Durch die Verwendung eines Bindestrichs werden Bereiche definiert. Die Angabe [a-s] bedeutet, dass jeglicher kleingeschriebene Buchstabe zwischen a und s gesucht werden soll. Der Ausdruck [1-3] bedeutet entsprechend, dass eine Zahl zwischen 1 und 3 gesucht wird.

Trefferanzahl festlegen: Die geschweifte Klammer {}

Um zu definieren, wie häufig sich ein Suchmuster maximal wiederholen darf, kann dies in geschweiften Klammern angegeben werden. Neben der exakten Angabe über {Zahl}, z. B. {2},

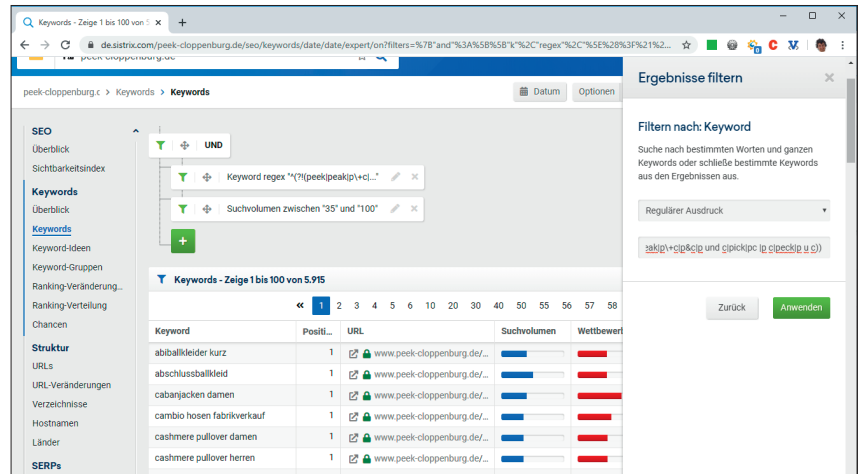


Abb.6: Durch den regulären Ausdruck werden Brand-Suchanfragen von Peek & Cloppenburg herausgefiltert

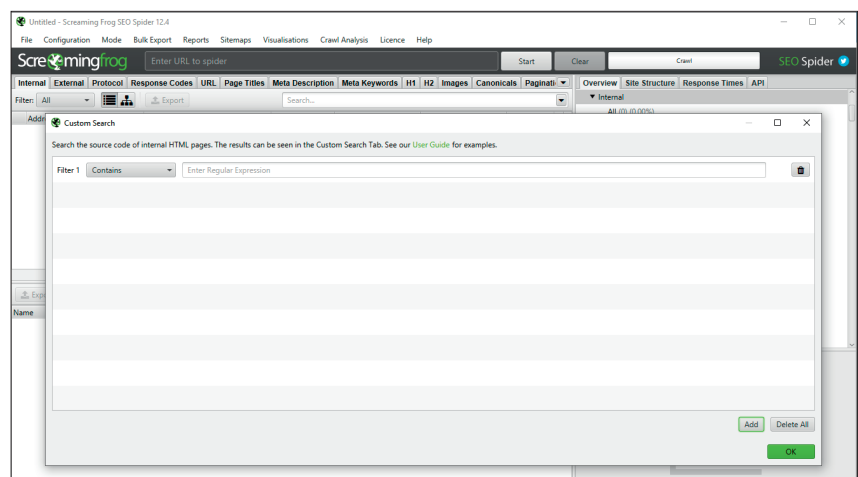


Abb.7: Sowohl für Suchen als auch zur Datenextraktion unterstützt der Screaming Frog RegEx. Unter Configuration => Custom kann mittels RegEx gesucht oder Text aus der Seite extrahiert werden. Zudem unterstützt die Suchfunktion für die einzelnen Tabs wie „Internal“ oder „External“ reguläre Ausdrücke (in der Abbildung als „Search...“ zu sehen)

kann zudem der Minimal- bzw. Maximalwert festgelegt werden. Während {1,3} für mindestens 1 und maximal 3 steht, stehen die Angaben {,2} für höchstens zweimal oder {2,} für mindestens zweimal. In Abbildung 3 sollen auf „ha“ mindestens ein bis maximal drei kleingeschriebene L folgen.

Nicht nur die Klammern oder der Punkt haben bei der Arbeit mit regulären Ausdrücken eine besondere Bedeutung. So gibt es Angaben, die z. B. die Suche nach Zahlen erleichtern.

- » \w findet sowohl klein- und großgeschriebene Buchstaben als auch Zahlen. Ausnahme bilden Buchstaben, die z. B. mit einem Akzent versehen sind wie é oder à.
- » \d findet Zahlen.

- » \s findet Leerzeichen, Tabulatoren und Ähnliches.
- » \b definiert, dass das davorstehende Zeichen das Ende der Zeichenkette darstellen muss.

Die Bedeutung dieser Angaben kann auf das Gegenteil umgekehrt werden, indem von Klein- auf Großbuchstaben gewechselt wird. Ein \D definiert folglich, dass keine Zahl gesucht wird.

Beispiel 1: Reguläre Ausdrücke in Google Analytics nutzen

Das beliebte Webanalyse-Tool von Google unterstützt die Suche mittels regulärer Ausdrücke. Dazu muss die erweiterte Filterung durch einen Klick auf „Erweitert“ neben dem Suchfeld

RegExr: Learn, Build, & Test RegEx

regexpr.com

Untitled Pattern Save (ctrl-s) New

Menu

- Pattern Settings
- My Patterns
- Cheatsheet
- RegEx Reference
- Community Patterns
- Help

RegExr is an online tool to **learn, build, & test** Regular Expressions (RegEx / RegExp).

- Supports **JavaScript & PHP/PCRE** RegEx.
- Results update in **real-time** as you type.
- **Roll over** a match or expression for details.
- **Save** & share expressions with others.
- Use **Tools** to explore your results.
- Full **RegEx Reference** with help & examples.
- **Undo & Redo** with ctrl-Z / Y in editors.
- Search for & rate **Community Patterns**.

Bring your team together with Slack, the collaboration hub for work.

Expression: `/^[0-9].*(tips|tutorial|how to)/gim`

Text: 4 matches (0.1ms)

Tools: Replace List Details Explain

Roll-over elements below to highlight in the Expression above. Click to open in Reference.

- Beginning. Matches the beginning of the string, or the beginning of a line if the multiline flag (m) is enabled.
- Character set. Match any character in the set.
 - 0-9 Range. Matches a character in the range "0" to "9" (char code 48 to 57). Case insensitive.

10 technische SEO-Ranking-Tipps | SEO-How-To-2020-SEO-Tutorial: 50-Tipps & Tricks für Anfänger
 27 wichtige SEO-Tipps für 2020: So kommen Sie auf Platz 1
 Die 10 wichtigsten SEO-Tipps, die Du kennen solltest
 12 Experten-SEO-Tipps für mehr organischen Traffic
 Eure 5 Fragen, unsere 5 Antworten: SEO-Tutorial-2020
 9 SEO-Tipps für Ihre Website

Abb.8: Da die Titel mit einer Zahl beginnen müssen, sind einzelne der extrahierten Seitentitel trotz des Auftretens der gewünschten Wörter keine Suchtreffer

Timme Cloud 2.0

Leistung satt!

NEU

Die Timme Cloud ist einfach bedienbar und individuell erweiterbar.

Ohne Vorkenntnisse.

Und wann immer Sie wollen!

Mehr Power.
Einfach zu bedienen.


TimmeHosting
nginx-Webhosting

SSD

99.99%
Uptime

HOSTING
MADE IN
GERMANY

timmehosting.de/cloud

aktiviert werden. Dieses finden Sie oberhalb der Datentabelle.

Durch den in Abbildung 4 zu sehenen RegEx werden die Daten anhand der URL eingeschränkt – nur Adressen, die mit /wissen oder /blog beginnen, sind in der Tabelle zu sehen.

Reguläre Ausdrücke kommen in Google Analytics regelmäßig beim Anlegen von Zielen zum Einsatz – oder um durch Filter für einzelne oder gar alle Datenansichten Veränderungen an den zu erfassenden Daten vorzunehmen. Eine Datenansicht ist dabei vereinfacht gesagt ein Ausschnitt der von Google Analytics erfassten Daten. So ist es sinnvoll, neben einer Ansicht „alle Daten“ (sprich keine eingestellten Datenausschlüsse über Filter) eine Datenansicht anzulegen, die die (feste) IP-Adresse des eigenen Unternehmens von der Datenerfassung ausschließt.

Ein Ausschluss der eigenen IP-Adresse ist sinnvoll, da sich das Surf-Verhalten der Mitarbeiter häufig fundamental von dem der Zielgruppe unterscheidet und jegliche Werte in der Webanalyse beeinflussen würde. Durch unterschiedliche Datenansichten würde es weiterhin die Daten aller Zugriffe geben – und zusätzlich eine Ansicht, die eben die Zugriffe aus dem eigenen Firmennetzwerk ausschließt.

Filter für einzelne Datenansichten können in der Verwaltung von Google Analytics vorgenommen werden. Sie müssen dabei wissen, dass durch die Filter ein- und ausgeschlossene oder z. B. durch Suchen und Ersetzen veränderte Daten nicht „rückgängig“ gemacht werden können. Behalten Sie deshalb eine ungefilterte Datenansicht!

Beispiel 2: In der Sistris-Toolbox mit regulären Ausdrücken arbeiten

Zu viele Daten können überfordern – deshalb bieten SEO-Tools wie Sistris Filter an. Neben den gängigen

Filtern wie „enthält“ oder „beginnt mit“ unterstützt Sistris zudem die Suche mittels RegEx – sowohl in der „einfachen“ als auch in der Expertensuche. Letztere können Sie z. B. bei „Keywords“ in der Kopfspalte unter Optionen aktivieren.

Ein Anwendungsfall: Wenn Keyword-Daten von bekannten Marken analysiert werden, dann tauchen regelmäßig Suchanfragen auf, die den Markennamen beinhalten – und für die eigene Website aufgrund der zu erwartenden niedrigen Klickrate eher uninteressant sein dürften. Diese Markenbegriffe lassen sich mittels RegEx herausfiltern – beispielhaft in Abbildung 6 für Peek & Cloppenburg zu sehen. Werfen Sie gerne zudem einen Blick auf das Tutorial von Sistris unter <https://www.sistris.de/tutorials/regulaere-ausdruecke-nutzen/>.

Beispiel 3: Den Screaming Frog mit regulären Ausdrücken weiter aufwerten

Der Screaming-Frog-Crawler ist ungemein mächtig und erlaubt sowohl RegEx-Angaben in den Suchfiltern als auch bei der Extraktion von Daten aus den gecrawlten Websites. Beide Optionen stehen nach einem Klick auf „Configuration“ => „Custom“ zur Verfügung. Während die Suche das Vorkommen des gesuchten Ausdrucks zählt, wird über die Extraktionsfunktion der Treffer ausgegeben.

Über die Standardkonfiguration erfasst der Screaming Frog bereits sehr viele Daten, die für jeden Online-Marketer interessant sind. Wer neben Seitentitel, Überschriften, strukturierten Daten & Co. noch mehr Infos aus dem Quelltext extrahieren möchte, der kann diese über die Extraktionsfunktion erhalten. Nachfolgend einige reguläre Ausdrücke, die interessante Daten aus der Website kopieren:

- » Google Analytics ID extrahieren: [,,'](UA-.*?)[,,']

- » Schema.org-Bewertungswert aus JSON-LD auslesen: „ratingValue“: „(.*)“
- » Schema.org-Publikationsdatum aus JSON-LD auslesen: [,,']datePublished[,,']: *[,,'](.*)[,,']
- » E-Mail-Adressen finden: [a-zA-Z0-9-_.]+@[a-zA-Z0-9-]+\.[a-zA-Z0-9-_.]+
- » Google Tag-Manager-ID kopieren: [,,'](GTM-.*?)[,,']

Mehr Beispiel-Codes sind im sehr lesenswerten Tutorial unter <https://www.screamingfrog.co.uk/web-scraping/> zu finden.

Übung macht den RegEx-Meister

Eine Antwort bin ich noch schuldig: Wie lassen sich Seitentitel finden, die mit einer Zahl beginnen und „Tipps“, „Tutorial“ oder „How To“ enthalten? Dafür ist der RegEx `^[0-9].*(tipps|tutorial|how to)` eine mögliche Lösung.

Sowohl für die Filterung als auch für die Extraktion von Daten können reguläre Ausdrücke ganz wunderbar eingesetzt werden. Die Grundlagen und einige Beispiele kennen Sie nun, doch erst durch die regelmäßige Anwendung wird das Schreiben von regulären Ausdrücken eine leichte(re) Übung.

Da sich besonders in der ersten Zeit Fehler einschleichen und zu viel extrahiert oder ersetzt wird, helfen Online-Tools wie <https://regex.com/> bei der Fehlersuche – und erklären, warum was gerade wie geschieht. In diesem Tool sind zudem weitere Tipps zu finden, die Ihre Fähigkeiten rund um reguläre Ausdrücke erweitern. ¶