



Mario Fischer, Patrick Lürwer

R-LEUCHTUNGEN!

Teil 6: Interne Ankertext-Dubletten finden

DER AUTOR



Mario Fischer ist Herausgeber und Chefredakteur der Website Boosting und seit der ersten Stunde des Webs von Optimierungsmöglichkeiten fasziniert. Er berät namhafte Unternehmen aller Größen und Branchen und lehrt im von ihm gegründeten Studiengang E-Commerce an der Hochschule für angewandte Wissenschaften in Würzburg.

In den Ausgaben bis 54 bis 58 der Website Boosting konnten Sie in der Serie „R4SEO“ von Patrick Lürwer nachvollziehen, wie man die kostenlose Software R verwendet, was sie leistet und wie man sie nutzbringend für die eigene Arbeit für SEO bzw. die Aufklärung im Online-Marketing einsetzen kann. R wurde ja ursprünglich für Statistik entwickelt. Wer das bisher als Entschuldigung verwendet hat, es deswegen links liegen zu lassen, dem sei versichert, dass er damit komplett falschliegt.

R kann für den Einsatz im Unternehmen, allen Bereichen voran „Online“, sehr viel mehr leisten, als statistische Berechnungen durchzuführen. Genau genommen ist es ein wirklich nützliches Helferlein bei allen Aufgaben im Umgang mit größeren Datenmengen, mit Daten, die man erst in eine gewisse Struktur bringen muss, und bei der automatischen oder halb automatischen Datenbeschaffung aus praktisch fast allen Quellen aus dem Web!

DER AUTOR



Patrick Lürwer ist Senior Analyst bei get:traction GmbH. Dort ist er für die Datenerfassung, -aufbereitung und -analyse zuständig. Sein tägliches Handwerkzeug sind R, Python und KNIME.

Für die interessierten Einsteiger, aber auch für alle, die nach der Serie von Patrick Lürwer „R-Blut“ gelectet haben, startete in der Ausgabe 62 die neue anwendungsorientierte Serie „R-Leuchtungen“. Sie werden in jeder Ausgabe erfahren, wie Sie ohne Programmierkenntnisse jeweils ein definiertes und in der Online-Praxis häufiger auftretendes Problem rund um das Thema Daten und Auswertungen lösen können. Und keine Sorge, die kleinen Hilfe-Tutorials nehmen Sie Schritt für Schritt an der Hand, sodass Sie auch als Neuling von der Power von R profitieren können. Was hält Sie also ab, das einfach mal auszuprobieren? Die einzelnen Schritte müssen Sie übrigens nicht im Detail verstanden haben. Um an die hilfreichen Daten für ein besseres Ranking zu kommen, müssen Sie im Prinzip nur nachmachen, was Sie hier beschrieben finden.

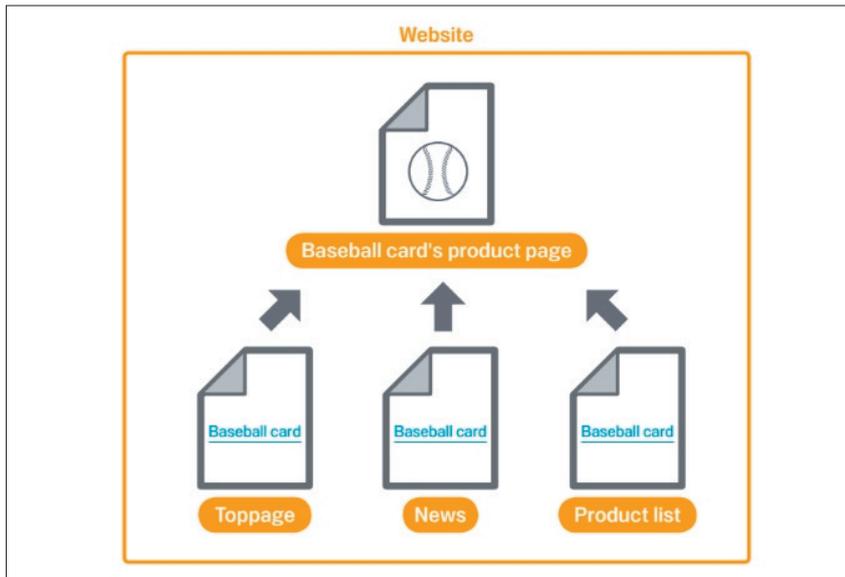


Abb. 1: Sehen Sie genau hin – die Linktexte sind identisch
(Quelle: Google, <http://einfach.st/seoguide1>)

Worum geht es diesmal?

Stellen Sie sich ein Hochhaus vor mit 100 Klingelschildern. Sie sollen ein Paket für Beatrice Hutschengolder ausliefern. Beim Blick auf die Klingelschilder stellen Sie fest, dass es 15 Klingeln mit dem Namen „Beatrice Hutschengolder“ gibt. Mist. Es mag Paketboten geben, die an dieser Stelle „unzustellbar“ in ihrem Handcomputer vermerken und schnell zurück in den Lieferwagen flüchten. Da Sie verantwortungsbewusst sind, kommt das für Sie nicht infrage. Aber was tun Sie jetzt? Wie gehen Sie vor?

In dieser Situation befinden sich sowohl Website-Besucher also auch die

Bots von Suchmaschinen, die auf Links stoßen. Da sich Menschen an den sichtbaren Linktexten (den sog. Ankertexten) für ihre Klickentscheidungen orientieren bzw. eine Erwartungshaltung durch den Ankertext entsteht, sollten die bekanntlich „sprechend“ sein. Also statt „hier klicken“ oder gar einem fast schon besucherverachtenden „more“ als sichtbaren Linktext verwendet man sicher besser vernünftige Beschreibungen, wohin man beim Klick genau kommt, wie etwa „Übersicht lila T-Shirts“ oder „T-Shirts vegan“. So weit, so gut. Der Ankertext sollte auf jeden Fall verbal relevant erklärend für die Zielseite sein.

So sieht ein Link im HTML-Code typischerweise aus:

```
<a href="https://www.meinedomain.de/seitel.htm">klicken Sie hier</a>
```

Und das erzeugt dann auf der Webseite diesen sicht- und klickbaren Text:

Klicken Sie hier

Kommen wir auf die Klingelschilder und Ihr Problem zurück. In welcher Wohnung finden Sie Beatrice Hutschengolder? Der Name ist selten, also ist es sehr unwahrscheinlich, dass in allen 15 Wohnungen tatsächlich je eine Frau mit

„Der sichtbare Text eines Links sollte möglichst erklärend und nicht allgemein formuliert sein.“

exakt gleichem Namen auf das Paket wartet. Welches wäre die beste Glocke für diesen Namen? Google sagt in einem Leitfaden zur Suchmaschinenoptimierung (<http://einfach.st/seoguide>): „Je besser dein Ankertext ist, desto einfacher können Nutzer die verlinkte Seite aufrufen und desto besser erkennt Google den Inhalt dieser Seite ...“, und gleich im nächsten Satz noch mal: „Mit einem geeigneten Ankertext ist der Inhalt der verlinkten Seiten für Nutzer und Suchmaschinen gut zu erkennen ...“

Google orientiert sich also an dem Ankertext, um den Inhalt der verlinkten Seite besser „zu erkennen“. Mit anderen Worten – und das ist in der Suchmaschinenoptimierung bereits seit über 20 Jahren bekannt –, werden die Wörter des sichtbaren Linktextes inhaltlich der Zielseite zugeschrieben und helfen dieser bei der Zuordnung zu Keywords bzw. am Ende eben auch zu einem besseren Ranking der Zielseite für diese – hoffentlich passenden – Texte.

Was hat das nun mit den 15 identischen Klingelschildern zu tun? Ganz einfach. Wenn man auf 15 unterschiedliche Zielseiten mit jeweils identischem Linktext verlinkt, wie etwa „lila T-Shirt“, dann kann diese Zuordnung „Ankertext“ -> „Keyword für die Zielseite“ natürlich nicht (maschinell) trennscharf durchgeführt werden. Die Suchmaschine kommt im schlimmsten Fall zu dem Schluss, dass es hier auf dieser Domain eben keine exakt beste Seite für lila T-Shirts gibt. Eine andere Domain hat auch eine Seite für lila T-Shirts, aber alle internen

INFO

Wer R noch nicht installiert hat, kann das schnell und einfach bewerkstelligen. Unter cran.r-project.org/bin/ findet man die aktuellen Versionen für Windows (32/64 Bit), MacOS (X) und Linux. Dort im Unterverzeichnis „base“ liegen dann die installierbaren Dateien. Nach der Installation holt man sich am besten gleich unmittelbar die Software „R-Studio“, die sich als eine Art Hülle um den Kern von R als grafische Benutzeroberfläche legt und viele Tools und Eingabehilfen zur Verfügung stellt. Die Bedienung von R wird mit R-Studio fast zum Kinderspiel. Man findet sie unter: rstudio.com/products/rstudio/download/

Links mit dem Text „lila T-Shirts“ zeigen nur dorthin und eben nicht noch auf viele andere Seiten der Domain. Und nehmen wir der Einfachheit halber an, alle anderen Rankingsignale zwischen den beiden Seiten Ihrer und der anderen Domain wären in etwa gleich – dann würde sich wohl jeder vernünftig Denkende für die eindeutige Seite der anderen Domain entscheiden, oder? Dort ist es so, dass das Klingelschild „lila T-Shirt“ immer mit der gleichen, eindeutigen Wohnung verbunden ist. Man ist hier einfach ein Stück sicherer, wenn man nur einen Klingelversuch hat. Man findet zwar bei Google keinen direkten Hinweis, dass man es vermeiden sollte, mit dem gleichen Ankertext auf unterschiedliche URLs zu verlinken, aber wenn man sich die Linktexte aus der erklärenden Grafik im Google SEO-Ratgeber genauer ansieht (Abbildung 1), erkennt man auch hier drei identische Linktexte („Baseball card“).

Zielführend für ein besseres Ranking ist also das „unique“ Verlinken. Das bedeutet, dass ein Linktext exakt so tatsächlich immer nur auf eine interne Zielseite zeigen sollte und nicht zu mehreren Seiten. Natürlich können auf diese Zielseite auch andere Linktexte zeigen, das wird von SEOs als nicht so kritisch angesehen. Es müssen also nicht alle Linktexte zu einer Seite identisch sein, aber Linktexte zu einer Seite sollten eben nicht auch auf andere Seiten zeigen. Das ist ein wichtiger Unterschied! Abbildung 2 zeigt dies nochmals schematisch auf. Auf die Seite „Lila T-Shirt“ gibt es zwei Ankertexte mit „lila T-Shirt“ und einen nur mit „T-Shirt“. Das ist o. k. Nicht o. k. ist allerdings, dass der Ankertext „lila T-Shirt“ noch mal verwendet wird und auf eine Seite mit einem lila Batikmuster führt. Dies gilt es zu vermeiden.

Doch wie findet man solche doppelten oder gleich mehrfachen Ankertexte? Noch dazu, wenn hier jahrelang durch verschiedene Editoren vielleicht ein

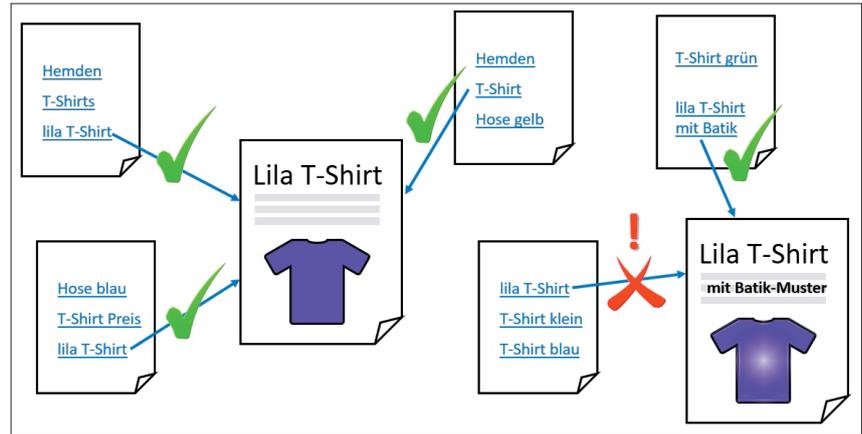


Abb. 2: Gleiche Ankertexte auf unterschiedliche Seiten innerhalb einer Domain sollte man unbedingt vermeiden

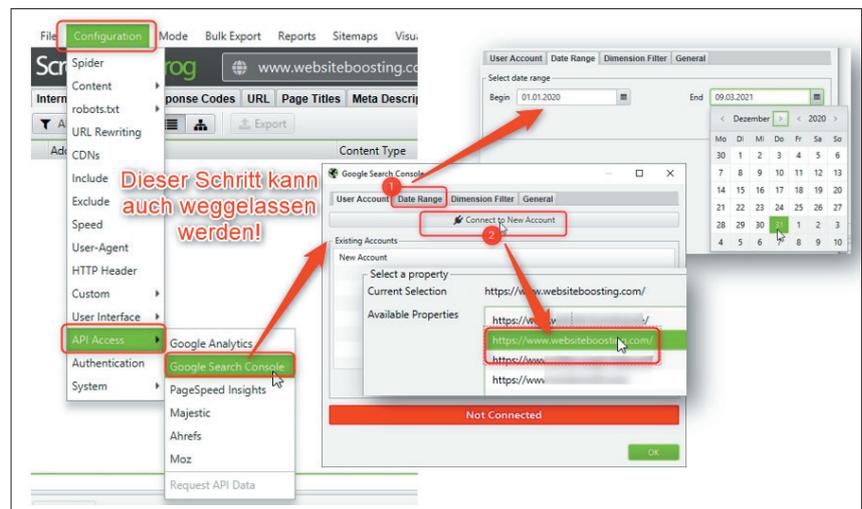


Abb. 3: Bonus: Die Daten der Google Search Console lassen sich bei Bedarf ganz einfach hinzufügen

gewisser Wildwuchs entstanden ist?

Ganz einfach: Mit unserem neuen R-Skript, das diesmal besonders leicht in der Anwendung ist.

Schritt 1: Die Linkdaten mit dem Screaming Frog abholen

Bei vielen SEOs zählt der Screaming Frog als Webcrawler zu den unverzichtbaren Werkzeugen und die meisten haben sicherlich eine lizenzierte Version des „Frosches“. Die Jahreslizenz schlägt beim aktuellen Umrechnungskurs mit etwa 170.- € zu Buche. Wer keine solche Lizenz hat, kann aber durchaus auch die kostenlose Version nutzen. Diese hat einige Einschränkungen, funktioniert aber für kleinere Websites bis zu 500 URLs bestens (siehe Tippkasten).

Es gibt grundsätzlich zwei Möglichkeiten, wie Sie vorgehen können.

1. Sie können für jede URL automatisch für eine von Ihnen definierte Zeitspanne (max. 16 Monate) die Klicks, die Impressions, die CTR (Click-Through-Rate) und die durchschnittliche Position bei Google in den Suchergebnissen abholen und anfügen lassen. Dies geht bequem über die im Frog eingebaute API, wie Abbildung 3 zeigt. Unter Configuration, API Access und Google Search Console öffnet sich ein Dialogfenster, in dem man über den Reiter „Data Range“ (Ziffer 1) das Zeitintervall für die Daten einstellen kann und sich anschließend über „Connect to New Account“ (Ziffer 2) ganz einfach mit Google verbindet. Der Browser öffnet sich dabei automatisch und Google fragt nach, ob und mit welchem Account man sich verbinden möchte. Das

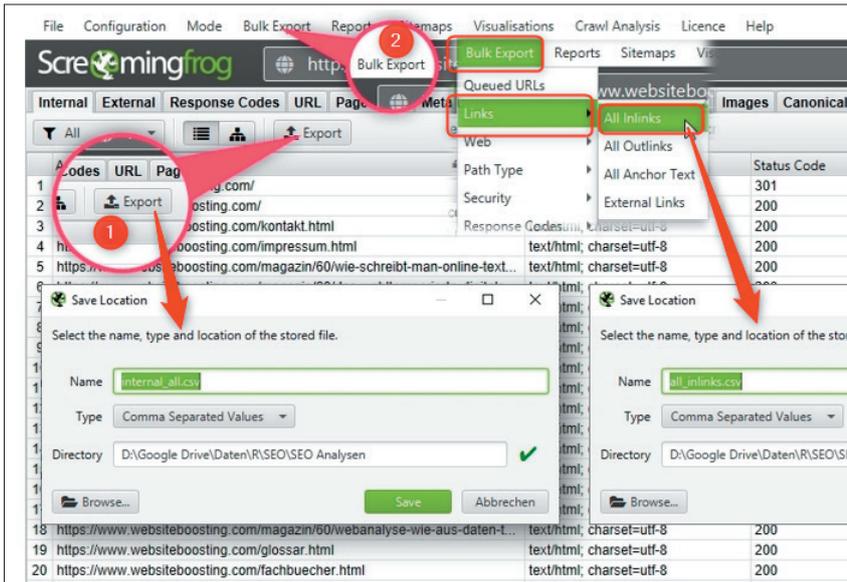


Abb. 4: Die benötigten Daten lassen sich aus dem Frog per Mausklick exportieren

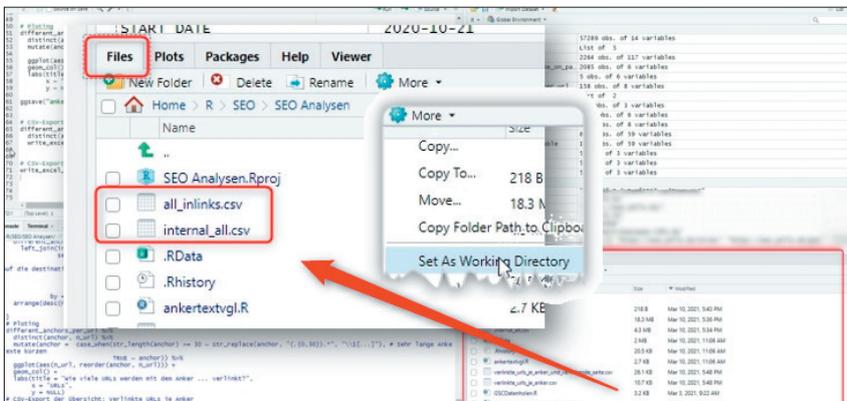


Abb. 5: So finden und ändern Sie das Arbeitsverzeichnis von R mit den Arbeitsdateien

TIPP

Screaming Frog

Wenn Sie noch keine Vollversion des Screaming Frog besitzen, können Sie sich auch eine funktionell beschränkte und kostenlos nutzbare Testversion unter www.screamingfrog.co.uk/seo-spider/ herunterladen. Diese ist zwar dann auf bis max. 500 URLs beschränkt, aber für kleine Domains ist das oft durchaus ausreichend. Und da man für die vorliegende Problemstellung nur HTML-Seiten braucht, kann man zum Sparen unter „Configuration“ und „Spider“ das Crawlen von Bildern, CSS, JavaScript-Files und SWF-Dateien getrost deaktivieren. Weiterhin beinhaltet die kostenlose Version auch keine API-Integration für die Google Search Console. Aber diese Daten brauchen Sie für eine reine Anker-text-Analyse auch nicht. Sie helfen nur, später zu sehen, welche der URLs wie viel Traffic über Google bekommen hat (eben durch das Anzapfen der Google-Daten über die API).

funktioniert natürlich nur, wenn man über einen solchen Account verfügt bzw. dieser auch angelegt wurde. Wer diesen tatsächlich noch nicht hat, der findet hierzu genügend Anleitungen im Web.

2. Sie verzichten auf die Datenergänzung von Google. Für die reine Anker-textanalyse wird sie nicht benötigt und stellt sozusagen nur einen Bonuswert am Ende der Analyse dar. Dazu später mehr. Und keine Sorge – wenn Sie erst mal den einfachen Weg zum Ausprobieren nehmen, können Sie später die Vorgehensweise unter 1. jederzeit problemlos nachholen.

Wenn Sie nur über die Free-Version des Frog verfügen, ist der Datenabruf über die API sowieso nicht möglich und Sie müssen den Weg 2 nehmen – aber

keine Sorge, alles im R-Skript funktioniert trotzdem.

Wenn Sie die API-Anbindung aktiviert haben oder diese weglassen bzw. nicht nutzen möchten, geben Sie einfach die Startadresse der Domain in die obere Eingabeleiste ein und drücken auf „Start“. Der Frog startet nun und crawlt jede erreichbare Seite durch. Rechts unten können Sie den fortlaufenden Status sehen – und auch am Ende die Zahl 100 %, wenn alles abgeschlossen ist.

Jetzt exportieren Sie einfach über den Button „Export“ das CSV-File „internal_all.csv“ (Abbildung 4, Ziffer 1) und über die obere Menüleiste unter „Bulk Export“, „Links“ und dann „All Inlinks“ die Datei „all_inlinks.csv“ (Ziffer 2). Am besten speichern Sie die beiden Dateien gleich in das Arbeitsverzeichnis in R. Wo das liegt, sehen Sie rechts unten in

R-Studio unter dem Reiter „Files“ (Abbildung 5). Oder Sie navigieren sich einfach dort per Klick in das Verzeichnis, in das Sie die beiden Dateien gespeichert haben. Anschließend klicken Sie unter „More“ beim hellblauen Zahnradsymbol auf „Set As Working Directory“. Dann ist dieses Verzeichnis bis auf Weiteres das Arbeitsverzeichnis. So weiß R dann, wo die zu verarbeitenden Dateien liegen.

Schritt 2: R bzw. R-Studio starten und das Skript einkopieren

Für das vorliegende Skript brauchen Sie keinerlei Eingaben oder Anpassungen. Sie können es daher einfach so, wie es ist, in R-Studio kopieren und am Stück durchlaufen lassen.

Starten Sie R-Studio. Wenn Sie in der Vergangenheit bereits unserer Serie gefolgt sind und R-Studio verwendet

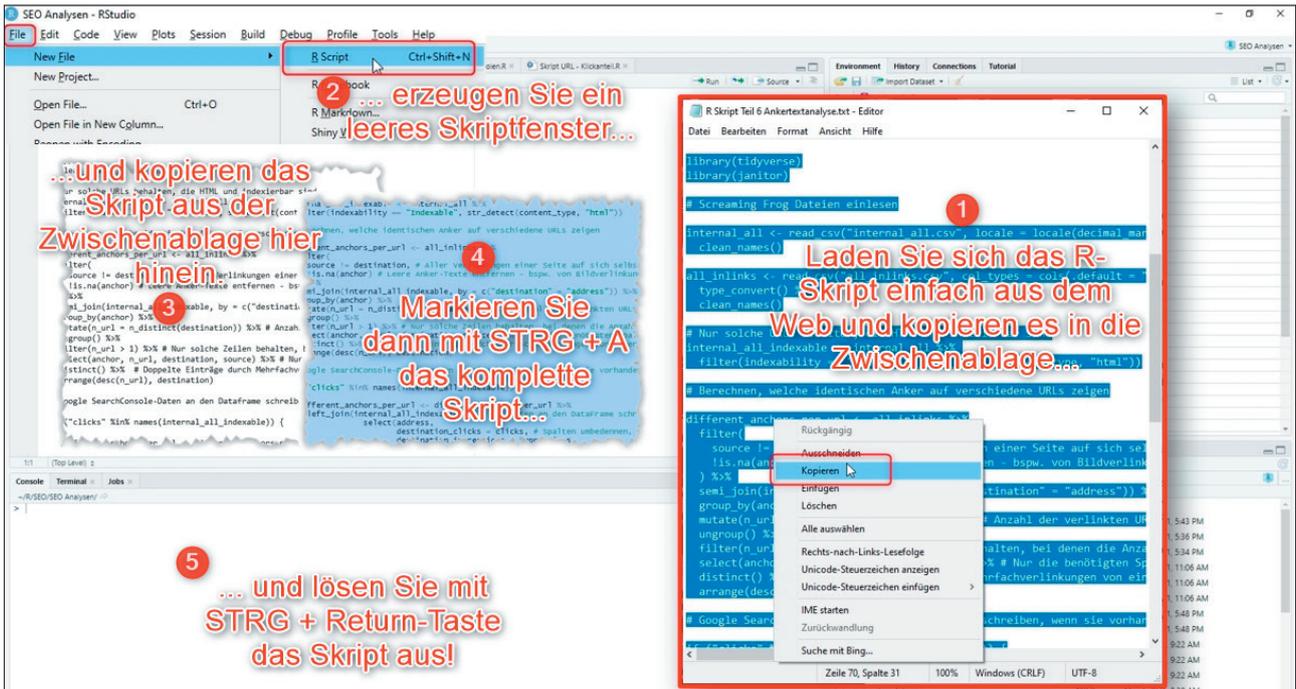


Abb. 6: Das Skript in R-Studio zu nutzen, geht ganz einfach in nur fünf Schritten

haben, können Sie die nachfolgenden Anweisungen ignorieren, weil die beiden benötigten Bibliotheken (Libraries) bereits installiert sind. Wenn Sie unsicher sind oder R-Studio das erste Mal verwenden, installieren Sie diese einfach noch mal bzw. erstmalig, das schadet überhaupt nichts.

Warum Sie das nicht im Code laufen lassen sollten (dort haben wir die beiden Zeilen mit einem # auskommentiert, d. h., sie werden mit dem vorangestellten # einfach ignoriert) hat den einfachen Grund, dass R dann jeweils neu starten möchte und das Skript (immer wieder) deswegen unterbrochen würde. Das ist sehr unpraktisch und lästig.

Schreiben Sie einfach links unten in das Konsolenfenster die beiden Anweisungen und bestätigen mit Return:

```
install.packages(„tidyverse“)
install.packages(„janitor“)
```

Diese Befehle installieren alles Nötige und nach dieser Installation startet R bzw. R-Studio neu. Diesen Installationsschritt müssen Sie dann später nicht mehr machen, er ist nur einmal erforderlich.

Ab jetzt wird es noch einfacher: Abbildung 6 zeigt Ihnen den kompletten

Ablauf in einem Bild:

1. Kopieren Sie den Code, den Sie unter <http://einfach.st/rcode6> zusammen mit allen bisherigen Skripten dieser Serie finden, in die Zwischenablage.
2. Oben in der Menüleiste von R-Studio unter „File“/„R-Skript“ erzeugen Sie ein neues leeres Skriptfenster im linken oberen Teil von R-Studio.
3. Dort hinein kopieren Sie das Skript aus der Zwischenablage.
4. Platzieren Sie den Cursor irgendwo im Skripttext und drücken Sie STRG und die Taste A – das markiert den gesamten Text und dieser ändert sich dann farblich (üblicherweise in Blau).
5. Drücken Sie STRG und die Return-Taste. Das löst das Abarbeiten des eben markierten Skripts aus.

Das war auch schon alles.

Da R sehr schnell arbeitet, kann das durchaus nur eine oder wenige Sekunden brauchen. Rechts unten ist eine Balkengrafik zur Verteilung aufgetaucht und im Arbeitsverzeichnis finden Sie drei neue Dateien mit den selbstsprechenden Namen:

- » verlinkte_urls_je_anker.csv
- » verlinkte_urls_je_anker_und_verlinkende_seite.csv
- » Rplot.png (die erzeugte Abbildung)

Schritt 3: Analyse der Ergebnisse

Die rechts unten erzeugte Balkengrafik zeigt Ihnen sofort auf einen Blick, ob Sie ein Ankertextproblem haben oder nicht (Abbildung 7 und Abbildung 8). Wenn Sie erkennen, dass wichtige Keywords hier mit einem Balken auftauchen, also auf mehr als eine URL zeigen, müssen Sie in die Detailanalyse und die beiden CSV-Dateien sehen. Und auch wenn eine Darstellung wie die in Abbildung 8 auf den ersten Blick erschreckt: Solange nur generische, unbedeutende Wörter wie „weiter“ oder „hier klicken“ mehrfach auftauchen, ist das zwar in den meisten Fällen nicht ideal, weil keine sprechenden Texte verwendet werden (siehe Erklärung am Anfang des Beitrags), aber nicht unbedingt „schädlich“ für das Ranking. Natürlich bleibt da Potenzial liegen, worüber Sie auf jeden Fall nachdenken sollten.

Ein Blick in die beiden Dateien hilft dann auch zu entscheiden, ob und an welchen Stellen man mit der Optimierung bzw. dem Aufräumen bei den Duplikaten startet. Jetzt und an dieser Stelle wird auch deutlich, warum das Anreichern jeder URL mit den Daten der Google Search Console so wertvoll ist.

In Abbildung 9 ist sofort zu erkennen, welche der 14 unterschiedlichen URLs, die mit dem gleichen Ankertext „lila T-Shirt“ angelinkt wurden, die besten Performancewerte bei Google haben. Diese Daten, Klicks, CTR und Rankingposition, helfen Ihnen zu entscheiden, welche URL die „richtige“ ist und auf welche die „Lila T-Shirt“-Ankertexte ausschließlich zeigen sollten. Verändern bzw. variieren Sie also alle Linktexte zu anderen URLs oder biegen Sie alle diese Links eben nur auf die eine ausgewählte um.

Fazit:

Geben Sie Google ein wenig Zeit, bis alle Veränderungen an den Linktexten und ggf. -zielen erkannt und neu bewertet wurden. Das kann (leider) durchaus einige Wochen bis Monate dauern. In der Regel werden Sie aber am Ende mit besseren Rankings für die ausgewählte Ziel-URL rechnen dürfen, weil Sie die „Signalqualität“ für das Ranking verbessert haben.

Alternativ können Sie die Idee der doppelten Ankertexterkennung auch rein in Excel durchführen. Dazu müssen Sie allerdings mehrfach filtern und eine gute Datensäuberung durchführen. Über eine Pivot-Tabelle lässt sich dann,

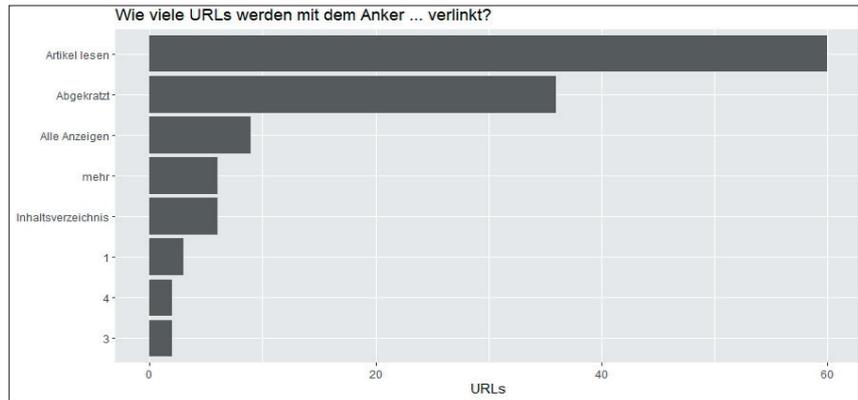


Abb. 7: Hier sind keine wichtigen Keywords als Ankertexte doppelt – Entwarnung ...

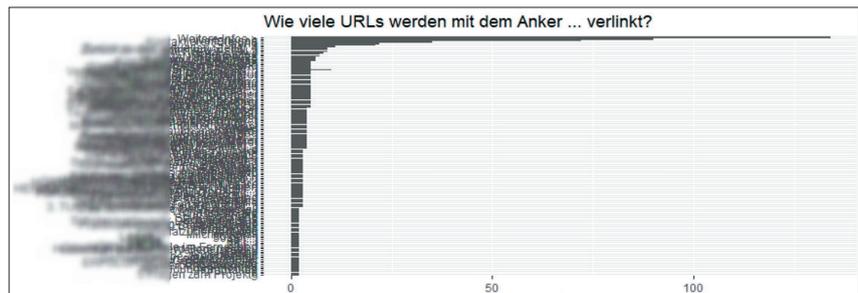


Abb. 8: ... aber hier ist optisch sofort zu erkennen, dass viele Ankertexte mehrfach verwendet wurden

nach der Spalte mit den Ankertexten und dem Wert „Anzahl“ für das Auftauchen mit unterschiedlichen URLs gefiltert, auch eine Liste mit doppelten Linktexten finden. Das ist nur deutlich komplizierter, dauert länger und ist auch fehleranfällig. Wem die einfache Nutzung von R-Studio noch immer etwas unheimlich vorkommt: Das Beispiel in diesem Beitrag ist wirklich sehr einfach nachzumachen. Ein idealer Ein-

stieg also und am Ende bringt es durch bessere Rankings auch noch mehr Traffic, Kundenanfragen oder bei Shops direkt mehr Umsatz. Sollten Sie bei der Analyse jede Menge URL-Pampe sehen mit vielen Parametern oder ähnlichen Strukturproblemen, wissen Sie zumindest schon mal, dass Sie ein ganz anderes Problem haben – und das sollten Sie dann natürlich zuerst angehen ... ¶

| anchor | n_url | destination | destination_clicks | destination_impressions | destination_ctr | destination_position |
|--------------|-------|-------------|--------------------|-------------------------|-----------------|----------------------|
| lila T-Shirt | 14 | https://www | 985 | 85.923 | 1,1% | 10,0 |
| lila T-Shirt | 14 | https://www | 11 | 3.672 | 0,3% | 32,7 |
| lila T-Shirt | 14 | https://www | 5 | 155 | 3,2% | 51,0 |
| lila T-Shirt | 14 | https://www | 10 | 1.951 | 0,5% | 31,8 |
| lila T-Shirt | 14 | https://www | 51 | 3.347 | 1,5% | 16,7 |
| lila T-Shirt | 14 | https://www | 22 | 2.031 | 1,1% | 39,0 |
| lila T-Shirt | 14 | https://www | 14 | 1.919 | 0,7% | 52,0 |
| lila T-Shirt | 14 | https://www | 949 | 54.310 | 1,8% | 20,7 |
| lila T-Shirt | 14 | https://www | 165 | 10.799 | 1,5% | 16,0 |
| lila T-Shirt | 14 | https://www | 46 | 4.025 | 1,1% | 28,2 |
| lila T-Shirt | 14 | https://www | 505 | 6.438 | 7,8% | 15,4 |
| lila T-Shirt | 14 | https://www | 10 | 1.419 | 0,7% | 47,3 |
| lila T-Shirt | 14 | https://www | 28 | 908 | 3,1% | 25,3 |
| T-Shirt | 8 | https://www | 3 | 327 | 0,9% | 5,2 |
| T-Shirt | 8 | https://www | 42 | 5.120 | 0,8% | 5,1 |
| T-Shirt | 8 | https://www | 4 | 173 | 2,3% | 13,3 |

Abb. 9: Man sieht sofort, welche (wichtigen) Ankertexte mehrfach verwendet wurden – und welche URL damit am besten Ranking und/oder Traffic bringt