Besser suchen & finden mit regulären Ausdrücken – RegEx im Google-Search-Console-Leistungsbericht

Besser suchen & finden mit regulären Ausdrücken: RegEx im Google-Search-Console-Leistungsbericht – websiteboosting.com

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…

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.

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.

Abbildung 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 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“, „pundc“, „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.

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 Hilfeartikel 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.

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

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:

p&c|peek|peak|cloppenburg|pundc

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.

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

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

Abbildung 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

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|ann|o|elche|as|obei|omit|oran|ohin|obei|eshalb|arum|ieso|orauf|orum|ovor|odurch|oher|eswegen|oraus)\s

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.

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

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.

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.

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

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.

Abbildung 6: Durch den Ausdruck ^(\w+)\s\w+\s\w+$ werden Drei-

Wort-Suchanfragen als Ergebnis geliefert

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.

Abbildung 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

An dieser 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.

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

ausgeblendet

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 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 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!

Backlinks – weniger ist mehr – Mit dem Google-Disavow-Tool zu langfristig besseren Rankings

Backlinks – weniger ist mehr. Mit dem Google-Disavow-Tool zu langfristig besseren Rankings – websiteboosting.com

Die Diskussion über die Wertigkeit von und die nötige Anzahl an Backlinks geht bereits seit vielen Jahren mit steigender Intensität durch die Branche. Die einen Experten sagen, Backlinks werden überbewertet und spielen fast keine Rolle mehr, andere, man könne nicht genug davon haben, und wieder…

Die Diskussion über die Wertigkeit von und die nötige Anzahl an Backlinks geht bereits seit vielen Jahren mit steigender Intensität durch die Branche. Die einen Experten sagen, Backlinks werden überbewertet und spielen fast keine Rolle mehr, andere, man könne nicht genug davon haben, und wieder andere erklären, man brauche gute Backlinks, um eine Chance zu bekommen, auf die erste Ergebnisseite zu gelangen – dort werde dann durch Tests von Google entschieden, ob man dortbleibt oder wieder abfällt. Mitarbeiter von Google selbst halten sich wie immer eher bedeckt zu dem Thema, betonen aber zwischen den Zeilen immer wieder, dass Backlinks schon noch einen deutlichen Einfluss hätten. Wenn Kaspar Szymanski über dieses Thema spricht, sollte man gut zuhören. Er war lange Jahre bei Google als Spamfighter in Dublin tätig und gehört daher zu der Personengruppe, die sich sehr gut auskennen dürfte …

Als Google 2012 das Disavow-Tool offiziell vorgestellt und damit effektiv die Gefahr, die für Webseiten-Rankings von vermeintlich unnatürlichen Backlinks ausgehen kann, eingeräumt hatte, wurde eine Wende vollzogen. Von da an waren Webseiten-Betreiber im Zugzwang, die Qualität ihrer Backlink-Profile aktiv zu überwachen. Backlinks, die potenziell Google-Richtlinien verletzten, mussten künftig mittels des neuen Disavow-Tools für ungültig erklärt werden. Doch dies war kein Neuanfang für alle Wettbewerber unter gleichen Vorbedingungen. Wer Backlink-Leichen im Keller hatte, musste sich deswegen gleich zu Beginn durchaus Sorgen machen. Selbst für jene Webseiten, die in der Vergangenheit ohne bewussten PageRank-Backlink-Aufbau ausgekommen waren, stellte die Einführung des Disavow-Tools eine nicht unbedeutende Veränderung dar. Schließlich wird jede noch so Google-konform gepflegte populäre Webseite mit der Zeit auch von minderwertigen oder gar von Spam-Backlinks geplagt. Auch diese mussten nun in regelmäßigen Abständen erfasst, evaluiert und für ungültig erklärt werden. Für zahlreiche Webseiten-Betreiber bedeutet diese Herausforderung auch einen erheblichen Ressourcenaufwand. Gleichzeitig stellt das Google-Disavow-Tool jedoch auch eine neue Wachstumschance dar. Vermag es doch, richtig eingesetzt, nicht nur unliebsame Google-Aufmerksamkeit oder gar potenzielle Bestrafungen zu vermeiden. Es hat viel mehr auch das Potenzial, Rankings langfristig steigen zu lassen.

> *„Der richtige Einsatz des Disavow-Tools kann das Ranking pushen!“*

Dieser Beitrag ist von der unmittelbaren Erfahrung des Autors sowohl mit dem Auferlegen wie auch dem Entfernen von Google-Strafen während seiner Tätigkeit bei Google Search geprägt. Es ist der Versuch, frei von Kommunikationsrichtlinien für Klarheit zu sorgen und Webseiten-Betreibern eine Orientierungshilfe anzubieten.

Ranking-Bremse Backlinks

Nicht alle Webseiten-Betreiber müssen Backlinks für ungültig erklären. Tatsächlich können die meisten Webseiten auf diesen Schritt verzichten, weil ihre Backlink-Profile zu unbedeutend sind oder weil es sich um Webseiten handelt, die nicht kommerziell betrieben werden und die niemals große Sichtbarkeit erreichen werden. Dazu gehören vor allem Hobby-Webseiten. Deren Betreiber können auf eine Laissez-faire-Strategie setzen, im absoluten Vertrauen auf die Unfehlbarkeit der Google-Algorithmen und darauf, dass diese Backlinks immer richtig klassifizieren können.

Im Gegensatz zu nicht kommerziell oder behördlich betriebenen Webseiten, wie etwa staatlichen Informationsplattformen, können die meisten Bezahldienste nicht auf Glück hoffen. Deren Betreiber sind angehalten, relevante SEO-Signale inklusive der Backlinks aktiv zu beeinflussen. Dabei nimmt das Disavow-Tool eine wichtige Stellung ein.

Abbildung 1: Wenn eine Backlink-bedingte Google-Strafe vorliegt, muss das Disavow-Tool verwendet werden

Es gibt allerdings auch Webseiten, die im Interesse der Google-Sichtbarkeit unbedingt das Disavow-Tool nutzen müssen. Das sind jene Seiten, die eine manuelle Maßnahme, besser bekannt als Google-Strafe, auferlegt bekamen. Solche Webseiten verlieren typischerweise zunächst abrupt und später kontinuierlich weiter an Google-Sichtbarkeit – so lange, bis das Problem gelöst wurde. Bei einer existierenden Google-Strafe, die die Webseite ohne Zweifel in den Google-Ergebnissen zurückhält, ist der Handlungsbedarf in Sachen Disavow-Tool akut.

Backlink-Analyse

Risiken, die von Backlinks ausgehen, können teilweise auch

mithilfe anderer Mittel minimiert werden, etwa durch gezielte Einträge in der Robots.txt-Datei oder mittels des NoFollow-Attributs. Das Disavow-Tool ist allerdings deutlich effektiver, insbesondere bei diversifizierten Backlink-Profilen oder bei Backlinks, über die der Webseiten-Betreiber keine Kontrolle hat.

> *„Der falsche Einsatz kann desaströse Folgen haben!“*

Allein diese Tatsache macht das Disavow-Tool zu einem unverzichtbaren Werkzeug im Online-Marketing. Zwar betrachtet Google die eingereichten Disavow-Dateien und deren Einträge als eine nicht bindende Empfehlung und behält sich das Recht vor, diese zu ignorieren. Doch das bedeutet nicht, dass das Disavow-Tool leichtfertig oder willkürlich eingesetzt werden sollte. Dessen Einsatz ohne eine Datenanalyse oder basierend auf Fehlinformation kann ungewollte, manchmal desaströse Folgen haben – z. B., wenn die Backlinks von Webseiten desselben Betreibers, die legitimerweise miteinander verlinkt sein können, unnötigerweise für ungültig erklärt werden. Eine erste Entscheidungshilfe, ob das Disavow-Tool überhaupt benutzt werden sollte, bietet die Google-Konsole selbst. Als Backlink-Datenquelle ist sie ebenso unverzichtbar wie eingeschränkt. Die 100.000 Beispiel-Backlinks stellen eher eine Stichprobe als das Gesamtbild dar. Gleichzeitig ist sie eine Referenzzahl und ein guter Wegweiser. Webseiten, deren Profil unter 100.000 Backlinks liegt, können auf den Einsatz des Disavow-Tools meistens verzichten. Jene Webseiten, die 100.000 und mehr PageRank-Backlinks angesammelt haben, sollten je nach Qualität der Backlinks das Disavow-Tool gezielt einsetzen.

Über das geschätzte Volumen von Backlinks kann man mit Zuhilfenahme von Online-Diensten wie Majestic häufig schnell und akkurat urteilen. Dabei sollte die Gesamtzahl von historischen und neuen Backlinks zusammen aufaddiert als ein Richtwert, jedoch nicht als ein absoluter Wert betrachtet

werden. Kein kommerzieller Dienst ist in der Lage, die exakte Anzahl aller Backlinks, die Google bekannt und damit relevant sind, zu ermitteln. Das ist jedoch auch nicht zwingend erforderlich, solange der Schwellenwert von 100.000 Backlinks berücksichtigt wird und der Backlink-Analyse eine für jede Webseite individuell ausreichend große Datenbasis zugrunde liegt.

Wichtig ist, dass Anbieter kommerzieller Backlink-Analyse-Werkzeuge, die häufig eine ebenfalls exzellente Datenquelle bieten, gleichzeitig fast immer an PBNs, also privaten Link-Netzwerken, scheitern und nicht in der Lage sind, diese eindeutig zu identifizieren. PBNs werden bewusst systematisch von Nicht-Googlebot-Crawlern abgeschirmt und deswegen oft von Drittanbietern nicht erfasst. Sie verdienen in diesem Zusammenhang jedoch nur eine Randnotiz, da die Idee hinter solchen Netzwerken konzeptionell falsch ist. Der PageRank-Backlink-Aufbau wird explizit für Google-Algorithmen betrieben. PBNs müssen folgerichtig für Google erfassbar sein, weswegen sie als eine klare Google-Richtlinien-Verletzung immer ein bedeutendes Risiko für das Ranking der Webseite darstellen. Die Erfahrung zeigt jedoch, dass die überwiegende Mehrheit aller Webseiten die von einer Backlink-Analyse profitieren könnten, nicht von PBNs betroffen sind.

Zu Beginn jeder Analyse werden, je nachdem, wie umfangreich ein Backlink-Profil ist, über einen geraumen Zeitraum möglichst viele Backlink-Daten gesammelt. Die Google-Search-Console stellt dabei nur eine von mehreren Quellen dar. Wie zuvor erwähnt, sind deren Datensätze begrenzt, was ihre Bedeutung für Webseiten mit 100 Millionen Backlinks oder mehr reduziert.

> *„Für eine Backlinkanalyse ist die Search Console nur eine von mehreren möglichen Quellen“*

Es ist zwar möglich, über die GSC einen höheren Output zu

generieren, indem neben der Domain-Property verschiedene URL-Kombinationen hinzugefügt werden, z. B. mit „www“, ohne „www“, „https“, „http“ sowie Kombinationen davon wie auch von Subdomains und Verzeichnissen. Der Ansatz ist jedoch bestenfalls eine Behelfslösung. Expertentools wie Ahrefs, Majestic, SEMrush und LinkResearchTools, die zu den populärsten auf dem Markt gehören, eignen sich hervorragend als Backlink-Datenquellen.

Eine möglichst breite Datenbasis ist die beste Ausgangslage für eine Backlink-Analyse. Die von Google kostenfrei bereitgestellten Backlink-Beispiele müssen deshalb um weitere Datensätze ergänzt werden. Die Bing-Webmaster-Konsole kann dabei ebenfalls gute Dienste leisten. Ähnlich wie die Google-Search-Console erlaubt das Bing-Webmaster-Tool einen begrenzten Einblick in das Backlink-Profil. Im Verlauf der Backlink-Analyse und der Untersuchung der Backlink-Qualität ist es unerlässlich, dass ein Crawler die Backlink-Daten, inklusive Anchor-Texte und Ziel bzw. Landingpages, erfasst. Die Backlink-Qualität als entscheidender Faktor ist schwer objektiv zu ermitteln. Sie hängt im Wesentlichen von den verwendeten Anchor-Texten, deren Häufigkeit im Auftreten, der Qualität der Inhalte sowie davon, welche Ziel-Landingpage verlinkt wird, ab. Es gibt zum jetzigen Zeitpunkt kein automatisiertes Verfahren, das diese kontextuelle Analyse verlässlich durchführen kann. Ausschließlich die Analyse eines Experten kann Aufschluss darüber geben, wie problematisch die Qualität eines Backlink-Profils tatsächlich ist.

> *„Die tatsächliche Qualität eines Backlinks ist schwer zu ermitteln*

Sobald eine repräsentative Datenmenge angesammelt wurde, kann die eigentliche Analyse beginnen. Große Datenvolumina können im Vorfeld gefiltert und dadurch reduziert werden. Zunächst müssen die Datensätze von Duplikaten bereinigt werden. Risikofreie Backlinks – etwa solche, die auf Landingpages

verweisen, die mittels robots.txt ausgesperrt sind, oder solche, die entweder mit dem Attribut rel="nofollow", "sponsored" oder "ugc" versehen sind – spielen für die Analyse keine Rolle. Sie haben keinen Einfluss auf die Risikoeinschätzung und können folglich ignoriert werden. Es sei denn, es handelt sich um bestätigte Spam-Backlinks, die erst kürzlich entfernt oder mit einem der Attribute versehen wurden. Denn Google hat die Veränderungen möglicherweise noch nicht realisiert. Qualitativ minderwertige Webseiten werden vom Googlebot tendenziell eher sporadisch aufgesucht, weswegen diese Backlinks nach wie vor ein Risiko darstellen können und unbedingt für ungültig erklärt werden müssen.

Dagegen können Backlinks innerhalb desselben Portfolios komplett ignoriert werden, wobei PageRank bzw. die verwendeten Anchor-Texte keine Rolle spielen. Da ein Webseiten-Betreiber berechtigterweise das gesamte Webseite-Portfolio verlinken kann, können alle Domains, die dazu gehören, von den Folgeschritten ausgeschlossen werden. Durch dieses Filtern im Vorlauf kann die anstehende Analyse relevanter Backlinks deutlich beschleunigt werden.

Webseiten mit vermeintlich hoher Domain Authority, einem Wert, der ausschließlich in den Köpfen mancher SEOs eine Rolle spielt, dürfen nicht von vornherein von der Analyse ausgeschlossen werden. Die so häufig zitierte Domain Authority spielt für Google keine Rolle. Erfahrungsgemäß verletzen Webseiten respektabler Organisationen bzw. bekannter Marken genauso häufig Googles Richtlinien wie alle anderen – ein Umstand, der Google im Laufe der Jahre nicht entgangen ist. Hat die Analyse der Backlink-Datensätze begonnen, liegt der Fokus auf deren Intention. Die zentrale Frage ist also, weshalb Backlinks gesetzt wurden, und nicht etwa, wer sie gesetzt hat oder wie alt sie sind. Für Google sind der Urheber und das Alter von Backlinks nicht vordergründig, weswegen eine detaillierte Dokumentation zu diesem Zeitpunkt nicht zwingend notwendig ist. Es genügt vollkommen, potenziell hochgradig

belastende Backlinks zu gruppieren, um diese effizient zu sichten und zu beurteilen. Dabei bestehen bei einigen Gruppen von vornherein kaum Zweifel, so z. B. bei Templates, also Vorlage-basiertem Spam oder Freehost-Spam mit unsinnigen Wortsalat-Inhalten.

> *„Google kann Backlinks auch heute noch nicht genau klassifizieren!“*

Während Filter bei der Vorsortierung durchaus eine Daseinsberechtigung haben, ist die eigentliche Analyse ein manueller Vorgang, der von speziellen Tools bestenfalls unterstützt werden kann. Keine auf dem Markt verfügbare automatisierte Lösung vermag dieses arbeitsintensive, detailorientierte Vorgehen zu ersetzen. Dies wird umso eindeutiger klar, wenn man bedenkt, dass selbst Google trotzt zwei Jahrzehnten an Erfahrung und nahezu grenzenloser Mittel immer noch nicht in der Lage ist, Backlinks genau zu klassifizieren. Wäre Google dazu in der Lage, würde das Disavow-Tool überflüssig. Und es gäbe auch keine manuellen Backlink-Strafen mehr.

Backlink-Signale

Im Verlauf der Analyse muss ein anderer gewichtiger Faktor berücksichtigt werden, nämlich die Anchor-Text-Verteilung. Dabei gibt es keine festen Schwellenwerte als Orientierungshilfen. Die Faustregel ist: Je mehr Top-Anchor-Texte für Produkte bzw. Dienstleistungen optimiert sind, desto höher ist die Wahrscheinlichkeit, dass ein PageRank-Linkaufbau durchgeführt wurde. Gleichzeitig steigt das Risiko, dass alte und neue Backlinks das Ranking der Webseite beeinträchtigen. Mehrere kommerziell verfügbare Dienstleister erlauben hierbei Einblicke, darunter Ahrefs und Majestic. Die Anchor-Text-Verteilung allein bietet keine ausreichende Entscheidungsgrundlage. In besonders wettbewerbsorientierten Branchen bietet sie aber oft einen ersten Eindruck, wenn unter

den Top-10-Anchor-Texten eine klare Tendenz erkennbar ist. Die Anchor-Text-Verteilung kann für die darauffolgende Untersuchung entscheidende Hinweise liefern.

Abbildung 2: Zahlreiche kommerzielle Top 10 der Anchor-Texte deuten oft auf suspekte Backlinks hin

Backlink-Typen

Für Google stellt eine Disavow-Datei eine Empfehlung des Webseiten-Betreibers dar. Ob diese vollständig oder teilweise befolgt wird, erfährt der Webseiten-Betreiber nicht. Allerdings bietet die Disavow-Datei einen zeitlich begrenzten Schutz. Wie lange dieser hält, hängt von künftigen Backlinks und deren Qualität ab. Aus diesem Grund sollte die Disavow-Datei als Teil einer Wartungsroutine mindestens einmal im Jahr auf der Basis aktueller Backlink-Daten überprüft und aktualisiert werden. Dabei ist es wichtig, alte und neue Einträge zu berücksichtigen. Einfach eine neue Datei mit URLs hochzuladen, führt zwingend dazu, dass alte Einträge unwiderruflich gelöscht werden, was das Risiko für eine Strafe erhöht.

Die eingereichte Disavow-Datei hat keinen Einfluss auf neue Besucher, die durch die Backlinks auf der Webseite landen. Gleichzeitig wird verhindert, dass Backlinks das Ranking negativ beeinflussen können.

> *„Die Disavow Datei sieht Google nur als Empfehlung an, bindend ist sie nicht“*

Es gibt keine Möglichkeit, exakt vorherzusagen, wie sich eine hochgeladene Disavow-Datei kurzfristig auf die Google-Sichtbarkeit auswirken wird. Drei unterschiedliche Szenarien sind dabei denkbar:

Die Rankings können unmittelbar

- stagnieren,
- sie können wachsen, wenn Backlink-Altlasten ihren Einfluss verlieren,
- oder sie fallen.

Eine Backlink-Analyse hat primär das Ziel, problematische Backlinks zu identifizieren. Es geht nicht darum, organische Rankings schnell zu erhöhen. Diesem Zweck dient einzig eine technische On- und Off-Page-Analyse. Langfristig trägt ein weitgehend Google-Richtlinien-konformes Backlink-Profil zum Ranking-Wachstum bei. Gleichzeitig reduziert es massiv die Gefahr einer Google-Strafe. Und: Der Linkaufbau kann und sollte nach wie vor weiter betrieben werden, allerdings mit Fokus auf die richtigen Prioritäten. Anstatt PageRank-Backlinks aufzubauen, was zu einer Abstrafung führt, ist es ratsam, Backlinks zwecks Konversion und im Hinblick auf das Erkennen neuer Inhalte (engl. „Content Discovery“) von Suchmaschinen-Bots sowie zur Bot-Priorisierung (engl. „Crawl Budget Prioritisation“) gezielt aufzubauen. Dieser Ansatz hat unter anderem auch den Vorteil, dass er nicht gegen Google-Richtlinien verstößt.

Risiko-Backlinks

Einzelne Backlinks stellen keine Bedrohung für eine Webseite dar, selbst wenn sie nicht Google-Richtlinien-konform sind. Es sind Backlinks als Daten betrachtet, die im Kontext einer Webseite und von deren individuellen Backlink-Profilen ein Risiko darstellen. Eine tiefgreifende Backlink-Analyse ist sowohl eine Kunst als auch eine Wissenschaft für sich. Wie komplex das Thema sein kann, wird vor allem dann deutlich, wenn man bedenkt, dass es Google trotz größter Anstrengung nicht geschafft hat, die Backlink-Klassifizierung gänzlich algorithmisch zu betreiben. Aus diesem Grund muss eine gründliche Backlink-Analyse manuell erfolgen. Dabei können allerdings einige Arten von Backlinks vorsortiert und als

entweder 100%ig harmlos oder als mit Sicherheit problematisch eingestuft werden.

> *„Einzelne schlechte Backlinks stellen kein Risiko dar“*

Sämtliche Backlinks, die keinen PageRank weitergeben, also mit dem NoFollow-Attribut versehen sind, können ignoriert werden. Das Gleiche trifft auf Backlinks zu, die auf Übergangsseiten verweisen, die mittels eines Robots.txt-Eintrags blockiert werden. Diesen mit Sicherheit harmlosen Backlinks gegenüber stehen PageRank-Backlinks, die über jeden Zweifel hinaus Google-Richtlinien verletzen und unbedingt in eine Disavow-Datei gehören. Zu jenen PageRank-Backlinks-Typen, die mit Sicherheit ein erhöhtes Risiko darstellen, gehören:

- **DoFollow-Affiliate-Backlinks**
- **DoFollow-PressRelease-Backlinks**
- **DoFollow-Themenfremde-Foren-Backlinks**
- **DoFollow-Guest-Post-Backlinks**
- **DoFollow-Directory-, -Startpagina- und -Verzeichnis-Backlinks**
- **DoFollow-Backlinks von gehackten Webseiten**

Alle verlinkenden Webseiten, die in eine der o. g. Kategorien fallen, müssen mit einem domain:-Operator-Eintrag (z. B. domain:example.com) in der Disavow-Datei berücksichtigt und damit für Google für ungültig erklärt werden.

Darüber hinaus gibt es andere Arten von Backlinks, die zwar nicht zu 100 %, jedoch mit hoher Wahrscheinlichkeit ein Risiko darstellen. Dazu gehören die folgenden Kategorien:

Risiko-Top-Level-Domains

Keine Top-Level-Domain besteht zu 100 % aus Spam. Bei einigen TLDs fehlt allerdings nicht mehr viel. Die Top-10-Statistik

des Spamhaus Projects bietet dazu aufschlussreiche Einblicke. Backlinks von Webseiten auf Domains wie .tk, .gq, .top, .ml oder .loan können zwar nicht pauschal als Spam abgetan werden. Doch kann man sie bei der Backlink-Analyse vorfiltern und gruppieren. In der Praxis erleichtern sich wiederholende Muster im Domainnamen oder in der URL-Struktur es oft, identische, automatisch erstellte Webseiten zu erkennen.

Gratis-Hosting-Anbieter

Bei näherem Hinsehen beinhalten nahezu alle Gratis-Hosting-Provider auch Spam-Webseiten. Einige Anbieter bestehen fast ausschließlich aus Spam – ein Umstand, der Google in der Vergangenheit dazu veranlasste, in extremen Fällen kollektive Strafen zu verhängen. Es ist nicht zielführend, präventiv alle Webseiten von kostenlosen Hostings in die Disavow-Datei aufzunehmen. Werden allerdings bei der Backlink-Analyse Spam-Webseiten erkannt, sollte hart durchgegriffen werden. In solchen Fällen ist es ratsam, den Anbieter des kostenlosen Hostings insgesamt zur Disavow-Datei hinzuzufügen, also z. B. domain:blogspot.com.

Backlinks von verfallenen Domains

Bei der Backlinks-Risiko-Analyse können einige Webseiten oder gar Domains leicht als potenziell problematisch eingestuft werden. Aufgegebene Domains, die früher mal legitime Webauftritte darstellten, jedoch im Laufe der Zeit aufgegeben wurden, gehören in diese Kategorie. Wenn zahlreiche solcher verfallenen Domains, die ausschließlich zwecks Linkaufbau Wiederverwendung finden, auf eine Webseite verlinken, ist das Risiko automatisch erhöht. Bei der Analyse und Verifizierung verfallener Domains bietet das Internet-Archive unschätzbare Dienste, da es den historischen Status von Domains bzw. Webseiten abbildet.

Abbildung 3: Das Internet-Archive (Wayback-Machine) erlaubt Einblicke in die Vergangenheit einer Webseite
Im Verlauf einer Backlink-Analyse fallen oft zahlreiche Backlinks an, die nicht eindeutig einer bestimmten Kategorie zugeordnet werden können. Bei der Risikoanalyse hilft es dann oft, die folgenden Fragen zu erwägen:

- Wird für den Backlink eine Gegenleistung erbracht?
- Kann der Backlink inklusive Anchor-Text kontrolliert und bestimmt werden?
- Wurde der Backlink automatisch generiert?

Wenn die Antwort auf mindestens eine dieser Fragen positiv ausfällt, stellt der Backlink wahrscheinlich ein erhöhtes Risiko dar.

Nicht zuletzt kann man gelegentlich auch die Frage aufwerfen, ob man den Backlink guten Gewissens einem direkten Konkurrenten oder aber einem Google-Mitarbeiter zeigen würde. Eine aufrichtige Antwort auf diese Frage hilft oft, die Spreu vom Weizen zu trennen.

Die Disavow-Datei

Sobald die Backlink-Analyse abgeschlossen ist und eine fast fertige Disavow-Datei vorliegt, gilt es nur noch, die Formatierung zu beachten. Während der Dateiname keine Rolle spielt, muss das Dateiformat des .txt-Dokuments UTF-8 oder 7-bit ASCII entsprechen. Google setzt für die Dateigröße eine Obergrenze von 2 MB und 100.000 URLs, was für die überwiegende Mehrheit von Webseiten vollkommen ausreicht. Es ist allerdings wichtig, alle Einträge mit dem domain:-Operator zu versehen.

Vor dem allerletzten entscheidenden Schritt ist es ratsam, die Disavow-Datei kostenlos und ohne Anmeldung auf ihre Formatierung hin überprüfen zu lassen. Das Online-Tool unter disavow.tools, betrieben vom ehemaligen Google-Ingenieur Fili

Wiese, ist in der Lage, in wenigen Minuten eine Disavow-Datei zu analysieren und gegebenenfalls Verbesserungsvorschläge zu machen.

Unmittelbar danach bleibt nur noch das Hochladen beim Google-Disavow-Tool unter: www.google.com/webmasters/tools/disavow-links-main. Während Google offiziell den Standpunkt vertritt, dass das Hochladen einer Disavow-Datei für eine Domain Property ausreichend ist, bleibt es dennoch ratsam, den Vorgang für alle verifizierten Einträge in der Google Search Console zu wiederholen, inklusive aller Kombinationen mit und ohne „www“, für sowohl „https“ als auch „http“ sowie etwaiger Subdomains und Verzeichnis-Einträge.

Und zuallerletzt gilt es noch, den Vorgang im Kalender vorzumerken, denn jede Disavow-Datei verliert mit der Zeit an Relevanz. Backlink-Profile wachsen und verändern sich mit der Zeit, weshalb eine Backlink-Analyse und eine Aktualisierung der Disavow-Datei regelmäßig durchgeführt werden müssen. Für die meisten Webseiten-Betreiber heißt das, dass die Backlink-Analyse ein fester Bestandteil des jährlichen SEO-Audits wird.

Machen Sie Ihren Doktor-Title! Neue Serie: SEO für Einsteiger – Teil 1

Machen Sie Ihren Doktor-Title! Neue

Serie: SEO für Einsteiger – Teil 1 – websiteboosting.com

Seit Anbeginn der Suchmaschinenoptimierung und sogar noch lange, bevor es diese Bezeichnung überhaupt gab, war der Title ein sehr wichtiges Element für Suchmaschinen. Es darf durchaus erstaunen, dass heute, mehr als 20 Jahre später, noch immer viele Websitebetreiber den Text des Titles eines…

Seit Anbeginn der Suchmaschinenoptimierung und sogar noch lange, bevor es diese Bezeichnung überhaupt gab, war der Title ein sehr wichtiges Element für Suchmaschinen. Es darf durchaus erstaunen, dass heute, mehr als 20 Jahre später, noch immer viele Websitebetreiber den Text des Titles eines Dokuments dem Zufall, mehr oder weniger untalentierten und/oder ungeschulten Editoren überlassen oder des Jammerns nicht müde werden, dass man doch nicht für jede Seite einen individuellen Title verfassen könne. Dabei sind die wenigen Wörter des Titles auch heute noch extrem relevant für das Ranking eines Dokuments! Und wer glaubt, bereits genügend über den Title zu wissen, den laden wir herzlich ein, dieses Wissen kritisch zu überprüfen und ggf. zumindest nachdenklich aufzustocken. Richtig gute SEOs kokettieren nicht selten damit, dass das Texten eines Titles „Chefsache“ sei. Nach dem Lesen dieses Beitrags werden Sie hoffentlich mit ihnen übereinstimmen, dass dies sooo weit hergeholt nicht ist. Aufgrund vieler Anfragen und Anregungen unserer Leser starten wir erneut eine kleine Serie mit SEO-Basics für Einsteiger und motiviert Fortgeschrittene. Im ersten Teil in dieser Ausgabe finden Sie alles Wichtige und viele Tipps zu einem der wichtigsten Bestandteile der Suchmaschinenoptimierung, dem Title.

Eines gleich vorweg: Auch der beste Title bringt Ihnen keine Toprankings, sofern die Gesamtkomposition an Signalen in Richtung Google nicht stimmig ist. Umgekehrt ist es extrem schwer, ein Seite-1-Ranking zu bekommen, wenn man im Title unpassenden oder nicht stimmigen Text verwendet.

Wow – Sie treffen heute den Menschen Ihres Lebens!

Stellen Sie sich doch bitte einmal vor, Sie säßen in einer Bar und entdeckten den Menschen (Frau, Mann, Divers) Ihres Lebens alleine an einem Tisch sitzend. Am liebsten würden Sie hingehen und ihn ansprechen. Das geht aber nicht, denn die Bar hat strenge Regeln. Gäste an Esstischen darf man nicht einfach so ansprechen bzw. belästigen. Und jetzt steht dieser Mensch auf und geht zur Toilette. Dabei muss er an Ihnen vorbei, direkt an der Bar. Sie sitzen zusammen mit neun weiteren Einzelpersonen an dieser Barzeile. Wenn der Mensch Ihres Lebens mit Ihnen Augenkontakt aufnimmt beim Vorbeigehen, haben Sie etwa zwei bis drei Sekunden Zeit, etwas zu sagen. Das ist Raum für sechs bis acht Worte. Natürlich überlegen Sie sich vorher ganz genau, was Sie sagen werden – oder etwa nicht? Sie würden kein Wort dem Zufall überlassen – oder etwa nicht? Und wenn Sie *wirklich* clever sind, testen Sie sogar, welche Worte die nötige Aufmerksamkeit erzeugen würden, damit dieser Mensch bei Ihnen stehen bleibt, lächelt und anfängt, mit Ihnen zu sprechen, Ihnen eine erste Chance auf „mehr“ gibt. Würden Sie sagen: „Na, auch hier?“ Oder: „Sektchen“? Würden Sie sagen: „Toilette – Toilette günstig besuchen, Toilette finden Sie hier?“ Gerade der letzte Satz kommt Ihnen völlig verblödet vor? Aber Moment – dieser Mensch wollte doch die Toilette aufsuchen. Was spräche dagegen, ihm genau diese Worte mehrmals um die Ohren zu werfen? O. k., eine ernste Antwort auf diese Frage können wir uns alle zusammen sparen.

Überträgt man die fiktiv beschriebene Situation auf die Kontaktanbahnung von Unternehmen mit potenziellen Interessenten bzw. Kunden, wird schnell klar, warum dieses Beispiel gewählt wurde. Jedem ist klar, dass Standardsprüche in Bars nur bei Menschen funktionieren würden, bei denen man das gar nicht wirklich will. Man muss sich Mühe geben, damit es bei denen klappt, mit denen man in Kontakt kommen möchte.

Abbildung 1: Bei vielen Unternehmen werden Titleinhalte offenbar extrem uninspiriert getextet. Schema F. Ich bin etwas Besonderes – daher texte ich ganz genauso wie alle anderen? Aber: Klick gefälligst mich?

Aber warum nur dringt diese einfache Erkenntnis nicht in die verantwortlichen Online-Abteilungen durch? Dieselben Menschen, die sich in einer Bar einen wirklich interessant wirkenden Spruch einfallen lassen würden, texten auf ihren Webseiten im Titel so Uninspirierendes wie: „Sandalen online kaufen/Domainname“? Ist es falsch verstandenes SEO? Fehlende Zeit? Oder die klassische Abwehrformel von Shops: „Wir haben [beliebige Zahl über 100 einsetzen] Artikel und können doch nicht für jeden einen eigenen Title texten!“?

> *„Sieht ein Suchender Ihren Title in einem Suchergebnis, ist das der ‚Moment of Truth‘ für Sie!“*

Was auch immer der Grund für falsch und/oder langweilige getextete Title ist: Sie sollten sich klarmachen, dass diese eine Zeile noch immer nicht nur wichtig für das Ranking ist, sondern auch für die Klickentscheidung von Suchenden. Mit einem guten Text bekommen Sie bei gleichem Ranking mehr Klicks und damit am Ende mehr Umsatz. Die Anzeige von Title und Description (siehe nächste Ausgabe) ist der „Moment of Truth“, also der Moment der Wahrheit. Zeigt Google sie zusammen mit neun anderen unbezahlten Treffern im sog. organischen Bereich in einem Suchergebnis an, bekommen sie ein bis zwei Sekunden Augenüberflug des Suchenden und er oder sie entscheidet sich spontan, ob es seine/ihre Zeit wert ist, Ihre Seite zu erkunden. Auf diese Einschätzung zahlt die Formulierung ganz entscheidend ein. Standard-Textgedöns hält sicherlich viele Suchende ab, wie sie in der Search-Console von Google als Diskrepanz zwischen guter Position, Impression (also wie oft in der Suche angezeigt) und Klicks (wie viele haben tatsächlich auf Ihren Title geklickt) leicht ersehen können.

Von vorne: Was ist der „Title“ überhaupt?

Laut Konvention ist <title> ein Element eines in HTML erzeugten Dokuments, also einer Webseite. Er zählt damit nicht zu den Meta-Tags wie z. B. der Description und dem völlig nutzlosen und trotzdem ständig befüllten „Keyword“-Meta-Tag.

Man findet den Title oben im <head> des Quelltexts (rechte Maustaste, „Quelltext anzeigen“ – oder ähnlich, je nach Browser). Ist er dort nicht sofort ersichtlich, drücken Sie STRG und die Taste f (Eselsbrücke: f wie finden). Jetzt ploppt ein kleines Suchfenster auf, in das man den Suchtext eintippt, also hier dann „title“ (ohne Anführungszeichen). So lässt sich bei umfangreichen Quelltexten gezielt finden, was man sucht.

Abbildung 2: Die Aussage in den ersten Dokumenten von Google zu SEO sind auch heute noch korrekt
Ein Beispiel: So sieht der Title im Quelltext einer HTML-Seite technisch korrekt aus:

```
<title>websiteboosting.com</title>
```

Im US-Patent Nr. 20070022110A1 von 2004 kann man nachlesen, dass die Wörter im Title gut als Rankingfaktor verwendet werden (können). Das Patent stammt allerdings nicht von Google, auch wenn es immer wieder im Zusammenhang mit den Rankingfaktoren von Google genannt wird.

Abbildung 3: Streng genommen falsch: Der Title ist ein fester Bestandteil einer HTML-Seite und eben kein Meta-Tag
Man kann sich den Title gut als den Namen eines Dokuments vorstellen oder auch als den kurzen, aber sprechenden Titel eines Buches, der auf dem Buchrücken steht. Sieht man ihn in der Buchhandlung im Regal zusammen mit anderen Büchern, entscheidet er in der Regel, ob wir das Buch in die Hand

nehmen und einen Blick hineinwerfen. Dieses Herausnehmen entspricht dem Klick auf Ihr Suchergebnis. Sind die Inhalte nicht gut, wird das Buch wieder zurückgestellt oder der Zurück-Button des Browsers geklickt. Der Title lockt also nur an, überzeugen muss dann aber eben auch der Content. Aber wenn schon der Title nicht lockt – wie soll dann der beste je gesehene Content verkauft werden können? Richtig, nie bzw. viel zu selten.

Für wen ist der Title und wo sieht man ihn überhaupt?

Auch wenn ein Seitenbesucher den Title also solches meist gar nicht bemerkt, weil er nur im Tab des Browsers erscheint oder als Textvorschlag für einen Favoriteneintrag bzw. ein Bookmark, schreiben Sie ihn trotzdem ausschließlich für Menschen. Oft ist man der irrigen Meinung, Titles müssten für Suchmaschinen und vor allem für Google geschrieben werden. Dies lässt völlig außer Acht, dass Google ausschließlich die Nutzer im Fokus hat und Dinge so gewichtet, wie diese Nutzer das präferieren. Es würde keinen Sinn machen, dies nach anderen Regeln zu werten. Ein Suchergebnis ganz oben mit einem schlecht lesbaren oder irrelevanten Title würde ja gar nicht geklickt werden und dort nur Platz beanspruchen. Verwerfen Sie daher am besten den Gedanken, für eine Maschine texten zu müssen, gleich wieder. Dieser entspringt einem falschen Verständnis, wie Google seine Kriterien fixiert. Nämlich (fast) nie nach eigenen Überlegungen, wie man es gerne hätte, sondern eben streng nach dem, was Google groß gemacht hat und groß hält: Was will der User? Betrachten Sie Google immer nur als eine Brücke zum Besucher, die hilft, den Kontakt herzustellen.

Den Title-Eintrag einer Webseite kann man wie erwähnt im Quelltext einsehen (Abbildung 4) und im Tab-Feld des Browsers (Abbildung 6). Die wichtigste Stelle ist allerdings sicherlich

die Darstellung in Suchergebnissen. In der Regel zeigt Google hier bis auf sehr seltene Ausnahmen den Title-Eintrag als Überschrift an (Abbildung 5). Ab und zu werden auch die Ankertexte eingehender (Back-)Links zu einer Seite als Title verwendet. Das passiert immer dann, wenn Google aus irgendeinem Grund dem vom Websitebetreiber hinterlegten Title misstraut oder z. B. jede Seite den gleichen Title trägt und dieser somit nicht aussagekräftig sein kann. Wird eine Seite als Link in Social-Media-Netzwerken geteilt, wird der Title als Überschrift verwendet, sofern für ein Netzwerk nichts speziell anderes hinterlegt wurde. Dies variiert jedoch von Plattform zu Plattform.

Abbildung 4: Den Title-Eintrag findet man im Quelltext oben im <head> einer Webseite

Abbildung 5: Im Suchergebnis wird der Title als blaue Überschrift angezeigt – in der Regel…

Abbildung 6. Der Title erscheint für einen Nutzer nur im Browser in der Fensterleiste oben bzw. in einem Tab als Bezeichnung

Sie wissen nicht, wo Sie den Title ändern/erzeugen können?

Leider ist bei Content-Management-Systemen nicht immer klar, welcher Eintrag im Backend dann tatsächlich in das Title-Element geschrieben wird bzw. welches Formularfeld das wirklich korrekte dafür ist. Dazu kommt erschwerend, dass bei Unternehmen oft nicht genügend Know-how vorhanden ist, um mit den Editor-Werkzeugen richtig umzugehen, die ihnen eine Webagentur im übertragenen Sinn auf den Tisch geworfen hat. Hier kann der eindeutige Rat nur lauten: Erkundigen Sie sich oder finden Sie selbst heraus, welcher Feldeintrag den Title „erzeugt“. Dazu sehen Sie einfach wie oben beschrieben im

Quelltext nach und suchen dann dieses Textmuster in Ihrem Editor. Gibt es mehrere Felder mit gleichem Inhalt, variieren Sie diese einfach und publizieren diese eine Seite neu. Rufen Sie diese dann im Browser auf und sehen sich im Quelltext an, wo genau die eben formulierten Texte „hingewandert“ sind. Identifizieren Sie den Text im <title>-Eintrag und ab jetzt wissen Sie genau, welches Editorfeld zum Title gehört.

Ein Beispiel. Der Quelltext zeigt z. B. zweimal den Eintrag „websiteboosting.com“ (Ausschnitt):

```
<meta name="generator" content="TYPO3 CMS">
<meta property="og:title" content="websiteboosting.com">
<title>websiteboosting.com</title>
<meta name="robots" content="index,follow">
```

Dann ändern Sie für diese Seite die generell infrage kommenden Felder um z. B. in „Hier ist Feld X“ und in das andere Feld „Hier ist Feld Y“. Nach der Publikation/Aktualisierung der Seite finden Sie im Quelltext dann:

```
<meta name="generator" content="TYPO3 CMS">
<meta property="og:title" content="Hier ist Feld X">
<title>Hier ist Feld Y</title>
<meta name="robots" content="index,follow">
```

Sie wissen jetzt, dass dort, wo Sie in Ihrem Editor „Hier ist Feld Y“ eingetragen haben, der eigentliche Title-Eintrag im Quelltext generiert wird. Voilà.

Wie gestaltet man einen guten Title?

Was empfiehlt eigentlich Google für die Gestaltung des Titles? Das ist nachzulesen unter http://einfach.st/gtip. Dort heißt es, man solle aussagekräftige Title verwenden, die den Inhalt der jeweiligen Seite kurz und knapp möglichst gut beschreiben. Irrelevante Wörter und Inhalte solle man auf jedem Fall vermeiden. Alles, was CMS oder Shopsysteme oft automatisch

einfüllen wie „Homepage“, „Seite 1“ oder „unbenannt“, sollte man auf jeden Fall mit sinnhaftem Text überschreiben. Auch das vor 20 Jahren übliche „Herzlich willkommen auf unserer Homepage“ wirkt bei vielen Suchenden wohl eher altmodisch und von vorgestern.

Abbildung 7: Welchen Eindruck erzeugen wohl solche Title beim Suchenden?

Abbildung 8: Leider zeigt Google ab und an auch völlig unpassende Ergebnisse im Ranking – hier sind die Title für die – dann unterbleibenden – Klicks extrem hilfreich

Abbildung 9: Die Texte auf Buchrücken wurden meist wohlüberlegt und Wort für Wort ausgewählt – so sollten auch Title gestaltet werden

> *„Es gibt Tests, die uns gezeigt haben, dass die Menschen geschriebene Titles bevorzugen.“ Gary Illyes, Google*

Ebenso wichtig ist es laut Google, dass jede Seite einen individuellen Titel bekommt. Die Verwendung gleicher Inhalte für mehrere oder gar alle Seiten einer Domain ist also unbedingt zu vermeiden.

Ein Beispiel von unserer eigenen Website:

<title>websiteboosting.com</title>

Finden Sie diesen Titel gut gestaltet für eine Startseite einer Domain? Nein, sicher nicht. Der wiederholt nur den Domainnamen und enthält daher keinerlei relevante Information, wenn er in den Suchergebnissen auftaucht. Noch nicht einmal „Startseite“ wurde als zusätzliche Information für Klickende hinterlegt. Wir leiden unter den gleichen Problemen wie so viele Unternehmen. Das wird halt einfach „erledigt“ und am besten ist es, wenn das CMS da gleich irgendwas automatisch reinschreibt, dann muss man sich erst gar nicht damit beschäftigen.

Wenn Sie an das Bar-Beispiel zu Anfang dieses Beitrags denken, wird noch mal klar, dass man sich tatsächlich überlegen muss: „Warum ich?“ Warum soll jemand unter vielen auf dieses Suchergebnis klicken. Natürlich kommen auch mit diesem Title viele Klicks – und zwar genau von denen, die das Magazin oder das frühere Buch „Website Boosting“ schon kennen. Aber jemand, der Know-how im Online-Bereich sucht? Dem „Website Boosting“ noch rein gar nicht sagt?

- Know-how satt für Online – Website-Boosting-Magazin?
- SEO, SEA, Web-Usability – Wissen für Ihren Erfolg?
- Die Zeitschrift für Ihren Weberfolg – Website Boosting?
- Seit über 10 Jahren feinstes Online-Know-how – guckst du?
- …

Im vorliegenden Fall müsste man noch nicht einmal überlegen oder gar testen. Alle Vorschläge wären sicher besser, als nur „websiteboosting.com“ auszuspielen. Natürlich muss bzw. sollte der Title SEO-relevante Wörter enthalten, das darf man nicht aus den Augen verlieren. Bei der Startseite ist allerdings in der Regel die Marke bzw. der Domainname selbst so stark, dass man mehr Freiheiten hat.

DEN richtigen, wahren Title gibt es in der Regel leider nicht. Bedenken Sie, dass eine URL natürlich auch für unterschiedliche Suchbegriffe und vor allem ggf. auch für verschiedene Suchabsichten ranken kann. Möchte jemand kaufen (Sandalen kaufen), wirkt ein Text für eine eher informationsgesteuerte Suche (welche Sandalen trägt man diesen Sommer) nicht optimal – und umgekehrt. Insofern muss man beim Texten oft Kompromisse eingehen, sofern man bisher nicht wirklich eine gute und vorausschauende Planung/Strukturierung beim Erstellen von Webseiten hatte. Insofern gibt es auch nicht DEN einen, alles einschließenden Ratschlag oder Tipp, der für alle gleich gilt.

Ihre Checkliste für die Gestaltung eines Title.

- Texten Sie immer **aussagekräftig und prägnant**.
- Vermeiden Sie **überflüssige Keywords und leere Füllwörter.**
- Vermeiden Sie **wiederholte Title, Title aus Textbausteinen oder eine automatisierte Erstellung.**
- Jede einzelne Seite sollte einen **einzigartigen Title** haben (Uniqueness)
- Länge: Bis max. 60 Zeichen (das sind 512 Pixel). Google schneidet zu lange Titles in den Suchergebnissen je nach Gerät ab, **Wichtiges** sollte daher immer ganz **vorne** stehen.
- Tipp: Das „**Abschneiden**“ kann natürlich an der richtigen Stelle auch **Neugierde** erzeugen.
- Großschreibung für ganze Wörter oder gar den gesamten Title ist schlechte Etikette, schwerer zu lesen bzw. zu erkennen und wirkt meist nach marktschreierischem „Brüllen“.
- Verwenden Sie wohlüberlegt Attribute je nach Produkt oder angebotener Dienstleitung wie z. B. „vollständig“, „umfassend“ oder „verständlich“. Auch die üblichen Aufzählungen a là „10 Tipps“ oder „Mit 2 Klicks zum …“ bringen mehr Klicks.
- Bei aktuellen Beiträgen ggf. den Monat und das Jahr im Title hinterlegen.
- Wenn die Marke *wirklich* bekannt ist: Unbedingt am Anfang nennen, um das Markenvertrauen in Klicks umzuwandeln.
- Vergleichen Sie die anderen Suchergebnisse, die bereits ranken. Seien Sie anders. Verwenden alle anderen bestimmte Begriffe wie z. B. „ultimativ“, verwenden Sie diese keinesfalls und suchen Sie nach besseren Alternativen.

- In der Regel funktionieren Emotionen gut, v. a. Ärger (vermeiden), Angst oder Neugierde.
- Bauen Sie sich ein eigenes Schema auf wie z. B. Keyword/Vorteil/Marke/Kategorie.

Wo kann ich das Aussehen meiner Title simulieren bzw. testen?

Im Web gibt es eine Menge „Simulatoren“, die für eine URL oder einen Text zeigen, wie er in den Suchergebnissen dargestellt werden würde, wie z. B.:

- app.sistrix.com/de/serp-snippet-generator
- searchwilderness.com/tools/pixel-length
- serpsim.com

Abbildung 10: Der SERP Snippet Generator von SISTRIX kann auch einfach „nur“ prüfen und gibt Hinweise (Quelle: Sistrix.com)

Abbildung 11: SERPsim.com zur Prüfung, wie der Title mobil oder auf dem Desktop (1) wirkt

Aber auch in vielen SEO-Tools wie z. B. in Ryte.com, ahrefs, sitebulb, Screaming Frog und vielen anderen kann man sich ansehen, wie eine einzelne URL im Suchergebnis mit Title und Description dargestellt würde. Als Hinweis bleibt zu erwähnen, dass fast alle dieser Tools ein klein wenig anderes rechnen, was die Länge bzw. den angezeigten Text angeht. Was die Länge angeht, sollten Sie also nur Anpassungen vornehmen, wenn Ihre Textlänge deutlich aus dem Rahmen fällt. Ist Ihr Title zu lang, prüfen Sie, ob wichtige Wörter möglichst weit vorne stehen und dann auch auf möglichst vielen Geräten gesehen werden. Ein paar Wörter „hinten“ schaden nicht. Und ein zu kleinliches Einkürzen kann Sie spätestens dann wieder einholen, wenn mehr und mehr über Sprachsuchen Ergebnisse von

Maschinen vorgelesen werden. Dann ist ein klein wenig mehr Info sicherlich durchaus hilfreich und Sie müssen dann nicht ergänzen, was sie jetzt gelöscht haben. Ist der Title zu kurz, stellt das oft eine verpasste Chance dar, noch mehr zu überzeugen und auch einige wichtige Wörter (für die Rankingalgorithmen) unterzubringen.

Was tue ich, wenn ich zu viele Seiten habe?

Immer dann, wenn etwas in zu großer Zahl auftritt, um sich mit allem einzeln im Detail zu beschäftigen, hilft das sog. Pareto-Prinzip. Es besagt vereinfacht ausgedrückt, dass man mit 20 % Aufwand 80 % des Erfolges steuert. Diese Faustformel trifft erstaunlich oft tatsächlich die Realität. Nicht selten sind bei Domains sogar nur 2 % der Seiten für mehr als 80 % der Besucher über Rankings verantwortlich.

Tipp 1: Filtern Sie!

Filtern Sie sich einfach über Ihr Tracking-Tool wie z. B. Google Analytics diese Seiten mit viel organischem Traffic von Google heraus und kontrollieren Sie die Titleeinträge auf Klick-Affinität. Stellen Sie sich die ehrliche Frage: Würde ich bei Suchen auf diesen Text klicken, wenn es nicht mein eigenes Unternehmen wäre? Macht der Text neugierig? Enthält er einen echten Klickanreiz? Sehen Sie sich die gesamten Suchergebnisse incl. Ihres Textes für eine entsprechende Suche bei Google an. Stechen Sie heraus oder gehen Sie im Textbrei der anderen mit unter? Das Wichtigste: Seien Sie ehrlich zu sich selbst!

Tipp 2: Holen Sie harte Daten!

Wenn Sie Titles wegen Tipp 1 ändern: Sehen Sie sich vor der Änderung die Klickmetriken in der Google Search Console an. Dort können Sie für jede URL und jedes Keyword, für das Sie

ranken, für einen Zeitraum von 16 Monaten sehen, ob und was sich verändert. Notieren Sie sich, für welche (wichtigen) Keywords die URLs ranken, bei denen Sie die Title optimiert haben. Nach den Änderungen sehen Sie nach einiger Zeit in den Suchergebnissen nach, ob Ihr neuer Title dort schon angezeigt wird. Wenn ja, warten Sie noch einige Wochen, damit aussagekräftige Zahlen zustande kommen, und kontrollieren Sie in der Search Console, ob sich die Klickrate verbessert hat. Behalten Sie aber im Auge, dass eine Veränderung der durchschnittlichen Position auch immer eine Änderung in der Klickrate zur Folge hat, weil Ergebnisse weiter unten oder weiter oben natürlich im Verhältnis fast automatisch mehr Klicks bekommen. Ist das Ranking aber im Schnitt in etwa gleich geblieben, können Sie ganz gut erkennen, was Ihre Änderung bewirkt hat.

Tipp 3: Lernen Sie daraus!

Stellen Sie bei der Befolgung von Tipp 2 fest, dass die Klickrate und damit die Anzahl der Besucher gestiegen ist, lernen Sie aus den eigenen Optimierungen! Was hat sehr gut funktioniert? Was gut, aber nicht so stark? Können Sie geänderte Textelemente identifizieren, die zu deutlich mehr Klicks geführt haben? Dann weiten Sie diese auch auf andere URLs aus. Damit gehen Sie in jedem Fall datengesteuert vor und nicht aus dem Bauch heraus – mit Letzterem liegt man in der Online-Branche leider oft komplett daneben.

Tipp 4: Kalkulieren und überzeugen Sie!

Versuchen Sie, Ihre Daten betriebswirtschaftlich zu verwenden. Ermitteln Sie das Delta der Mehrbesuche auf Ihrer Website. Wenn Sie Ihre Conversion-Rate kennen, also das Verhältnis von Besuchern zu Käufern (oder was Ihren Zielkriterien entspricht), können Sie diesen zusätzlichen Umsatzwert zumindest rechnerisch grob beziffern. Ein einfaches Beispiel: Sie haben eine Conversion-Rate von 5 %. Das heißt, von 100

Besuchern kaufen fünf etwas ein (oder schicken ein Kontaktformular ab, rufen Sie an, abonnieren den Newsletter etc. – je nach Ziel). Der durchschnittliche Warenkorb liegt bei 200 €. Also haben Sie monatlich bei 100 Besuchern 1.000 € Umsatz. Die Klickrate auf Ihr Suchergebnis lag vor der Optimierung laut Search Console bei 10 %, nach der Optimierung bei 20 %. Bei gleichen Rankings haben Sie also Ihre Besucherzahl von Google verdoppelt. Lassen Sie uns annehmen, dass es vorher 1.000 Besucher pro Monat und damit 10.000 € Umsatz waren. Jetzt sind es 2.000 Besucher und 20.000 € Umsatz.

Geht man von einer Gewinnmarge von 20 % aus, haben Sie mit den Änderungen 2.000 € mehr Gewinn gemacht (20 % von 10.000 € = 2.000 € und von 20.000 € wären das 4.000 €, also 2.000 € mehr).

Jetzt haben Sie etwas zum Kalkulieren in der Hand. Pro Monat 2.000 € sind 24.000 € mehr Gewinn pro Jahr. Stellen Sie diesem Zusatzgewinn (oder dem Zusatzumsatz) die Kosten der Änderungen gegenüber. Liegen diese deutlich darunter – und das tun sie in den allermeisten Fällen –, stellt sich die Frage nach der Sinnhaftigkeit weiterer Optimierungen nicht mehr. Sie ist nachgewiesen. Überlegen Sie, wer Optimierungen an weiteren Seiten vornehmen kann, den Gewinn-/Umsatzhebel kennen Sie jetzt und können auch kritischen Vorgesetzten (bringt das wirklich was?) gegenüber richtig und mit Geld argumentieren. Wichtig: Sie kommen jetzt nicht mehr als Bittsteller nach Budget daher, die immer nur Geld wollen für undefinierte Tätigkeiten. Sie können vorrechnen, was ein investierter Euro pro Jahr an Return-on-Investment bringt. Der Chef, der hier noch Nein sagt, ist aller Voraussicht nach falsch auf seinem Posten…

Tipp 5: Rankingplus – für Experten!

Die Erklärung der genauen Umsetzung würde hier den Rahmen leider sprengen, aber wer einigermaßen gut mit Tools umgehen

kann, für den mag die Idee hinter dem folgenden Tipp sicherlich recht hilfreich sein. Besorgen Sie sich eine Liste mit Ihren Rankings nach Keyword und URL aus dem SEO-Tool Ihrer Wahl. Bei SISTRIX finden Sie diese z. B. in der Unterrubrik „Keywords“. Rechts oben ist dann der „Export“-Button für den Download einer Excel-Liste. Filtern/sortieren Sie die Liste so nach Rankingpositionen, dass die guten Rankings (z. B. Position 1–5) nicht angezeigt werden. Diese brauchen Sie ja nicht optimieren. Jetzt müssten Sie für jede URL den Title holen. Das geht sehr geschmeidig z. B. mit SEO-Tools für Excel. Die Formel =@HtmlTitle(B1) holt den Title einer URL, die in der Zelle B1 steht. Wenn Sie diese Formel nach unten kopieren, steht für jede URL der Titletext in einer Spalte. Anschließend prüfen Sie, z. B. mit der Funktion =SUCHEN(A1;B1), ob das rankende Keyword in Zelle A1 im Text des Titles in Zelle B1 enthalten ist. Die Funktion =SUCHEN hat gegenüber =FINDEN den Vorteil, dass Groß-/Kleinschreibung ignoriert wird. Trotzdem muss man noch auf Leerzeichen und Bindestriche achten, aber mit ein paar Excel-Kenntnissen bekommt man das auch recht einfach in den Griff. Wer SEO-Tools für Excel nicht zur Verfügung hat, kann sich den Title auch mit anderen Tools wie z. B. dem Screaming Frog holen und via =SVERWEIS die erzeugte Datei und die aus dem Rankingtool matchen. Das Ergebnis ist eine Liste, die zeigt, welche Rankings schon gut, aber noch nicht top sind und bei denen das Keyword eben (noch) nicht im Title vorhanden ist. Ergänzt man das Keyword dort im Title, kann man in der Regel einen kleinen Sprung in den Rankings nach oben erzeugen. Gerade bei Ergebnissen auf Position 11 bis 15 kann ein Sprung auf Seite eins durchaus „plötzlich“ Traffic bringen, der bekanntlich von Seite zwei fast nie zu verzeichnen ist. Eine einfache Möglichkeit also, „low hanging fruits“ zu ernten.

Tipp

Die Inhalte dieses Beitrags und weitergehende Infos finden sie auch bei Website Boosting TV unter http://einfach.st/wsbtv5

erklärt. Abonnieren Sie uns doch einfach kostenlos auf YouTube – dann verpassen Sie ab jetzt keinen Videobeitrag mehr!

Fazit

Sie haben eine genaue Anweisung erwartet, welches Wort Sie wie und wo ändern sollten? Das ist aus verschiedenen Gründen nicht sinnvoll, allein schon deswegen, weil Sie ja „anders“ als die anderen sein sollten, hervorstechen sollen. Je mehr Sie nach Schema F vorgehen und je weniger Sie ihre eigene Kreativität für die Spezifika Ihrer Branche und Angebote nutzen, desto mehr entfernen Sie sich von einem guten Ziel. Denken Sie für Ihre wichtigsten Seiten an die Bar. Der Suchende geht vorbei, fixiert alle zehn nacheinander und Sie haben nur diese eine, kurze Chance, die nötige Aufmerksamkeit zu bekommen. Wenn Sie das verbummeln, bekommen Sie definitiv keine weitere Möglichkeit zu zeigen, dass Sie gut sind. Behalten Sie daher diesen wichtigen Moment of Truth immer im Kopf!

Wie Sie verkaufen, auch wenn Sie teurer sind

Wie Sie verkaufen, auch wenn Sie teurer sind – websiteboosting.com

Zugegeben, aktuell geht es der Digitalwirtschaft sehr gut. Zum einen wissen viele Kunden um die Wichtigkeit ihres digitalen Auftritts Bescheid, zum anderen unterstützt der Staat zusätzlich mit zahlreichen Förderprogrammen Investitionen – und damit auch so manche Agentur. Doch egal, wie die…

Zugegeben, aktuell geht es der Digitalwirtschaft sehr gut. Zum einen wissen viele Kunden um die Wichtigkeit ihres digitalen Auftritts Bescheid, zum anderen unterstützt der Staat zusätzlich mit zahlreichen Förderprogrammen Investitionen – und damit auch so manche Agentur. Doch egal, wie die aktuellen Konstellationen sind: Es wird immer Agenturen geben, die teurer als andere sind. Vielleicht gehören auch Sie zu den Höherpreisigen. Dann ist die entscheidende Frage: Wie erfolgreich schließen Sie Aufträge ab, wenn Sie nicht der billigste Anbieter sind? Oder knicken Sie gar am Ende der Verhandlung ein und geben zahlreiche Leistungen gratis, weil Sie nicht in der Lage sind, Ihren Preis auf Augenhöhe zu erklären?

So manche Gründer orientieren sich bei der Preisfindung an ihren Mitbewerbern: Was nehmen diese pro Stunde? Wie sehen deren Angebote aus? Im Ergebnis ist dann deren Angebot für Außenstehende oft austauschbar. Denn wenn Webseiten, Imagebroschüren und auch die Mitarbeiter alle die gleichen Phrasen dreschen – woran soll sich dann ein Kunde orientieren? Erst recht dann, wenn ein potenzieller Abnehmer wenig Ahnung von der Materie hat und zwangsläufig einen objektiven Vergleich zwischen beispielsweise drei Agenturen nicht machen kann, wird sich dieser oft für den günstigsten entscheiden müssen. Denn wie soll ein Kunde seinen eigenen Kollegen erklären, dass er nicht das billigste Angebot gekauft hat, sondern das teurere für 10.000 Euro mehr, wenn er das Angebot und das Konzept dahinter nicht wirklich versteht? Sie können davon ausgehen, dass manche Kunden bei Ihrem günstigeren Mitbewerber kaufen, obwohl Sie eigentlich den Auftrag wirklich verdient hätten, weil Sie ihm nicht ausreichend verdeutlichen konnten, dass Sie Ihren Preis wirklich(!) wert sind.

Am Anfang steht die Bedarfsanalyse

Das Wichtigste im Verkaufsgespräch, die Bedarfsanalyse, wird selbst von Profi-Verkäufern oft unterschätzt. Eine schlechte

Bedarfsanalyse rächt sich spätestens bei der Preisverhandlung. Denn wie wollen Sie argumentieren, wenn Sie nicht wissen, was Ihrem Kunden wichtig ist? Bei einer guten Bedarfsanalyse laufen Sie bildlich gesprochen im Kopf des Anfragenden herum, um diesen vollumfänglich zu verstehen. Sie wissen nach dieser Analyse, was Ihr Kunde will und nicht will. Sie wissen danach, woran Ihr Kunde festmacht, dass Sie seinen Auftrag gut erledigt haben – und dass er Ihren Preis respektiert und akzeptiert. Aufgrund der Antworten können Sie dann auch für sich entscheiden, ob Sie diesen Kunden überhaupt haben wollen – oder nicht. Oder etwas spitzer formuliert: Sie müssen nach der Bedarfsanalyse ganz klar für sich entscheiden, ob Sie es verantworten können, den Auftrag zu übernehmen. Denn wenn zwei Personen einen Vertrag unterschreiben, dann müssen beide Seiten auch liefern.
Teilen Sie die Bedarfsanalyse auf wenigstens(!) zwei Termine auf. Den ersten Termin führen Sie online oder telefonisch durch. Finden Sie bei diesem Gespräch Antworten auf mindestens folgende fünf Fragen:

- Was haben Sie genau vor?
- Weshalb?
- Ab wann soll das Ganze fertig sein?
- Welche Erfahrungen haben Sie bisher mit Agenturen gemacht?
- Wie sind Ihre finanziellen Vorstellungen?

Hauptziel dieses Erstgespräches ist es, dass Sie danach ganz klar sagen können: „Ja, es macht Sinn, weitere Zeit in das Projekt zu investieren“, oder: „Dafür sind wir nicht die Richtigen. Hier können wir es nicht verantworten, den Auftrag zu übernehmen.“ Im letzteren Fall sollten Sie den Kunden nicht im Regen stehen lassen, sondern im Idealfall an eine andere Agentur weiterempfehlen.

Kunden wollen Sicherheit

Aber nicht nur Sie müssen den Anfragenden „abklopfen“, sondern Sie müssen sich auch richtig inszenieren. Und zwar so, dass im zeitlichen Verlauf der Verkaufsabschluss zu Ihren kalkulierten Preisen das Ergebnis ist. Es bringt wenig, wenn Sie mit Interessenten langweilige Logosammlungen Ihrer zufriedenen Kunden durchgehen. Denn nur weil Sie Logos von Kunden veröffentlichen, bedeutet das ja noch lange nicht, dass Sie diese Kunden auch wirklich begeistert haben. Darüber hinaus sagt ein Logo auch recht wenig über die Größe des Projekts aus. Fragen Sie sich besser, welche Sorgen und Ängste potenzielle Kunden haben:

- Haben die mich wirklich verstanden?
- Können die das wirklich?
- Wie verhalten die sich, wenn Probleme auftreten?
- Gibt es unangenehme Überraschungen (finanzieller oder zeitlicher Art)?
- Was passiert, wenn ich während der Zusammenarbeit merke, dass ich mich für den falschen Anbieter entschieden habe?
- Sollte ich jetzt kaufen – oder bekomme ich diese Leistung auch irgendwo anders billiger oder besser?
- Wie viel „Luft“ ist noch in den Preisen?
- Was sagen die anderen Agenturen, die ich auch angefragt habe, wenn ich mich jetzt gegen diese entscheide?
- Wie werden meine Kollegen reagieren, wenn ich die falsche Agentur beauftrage?
- Wäre es besser, gar keine Agentur zu beauftragen – und alles so zu lassen, wie es jetzt ist?

Auf diese Fragen, die sich nahezu jeder potenzielle Kunde mehr oder weniger bewusst stellt, geben die meisten Anbieter von sich aus kaum gute Antworten. Zwangsläufig verlaufen so manche Anfragen im Sand. Denn wer hat noch nicht beim Nachfassen

Sätze gehört wie: „Wir lassen erst mal alles so, wie es ist“, oder: „Wir haben jetzt andere Themen auf der Agenda. Melden Sie sich gerne in einem halben Jahr wieder, dann haben wir dafür wieder den Kopf frei“?

> *Der Preis ist egal, wenn die Gegenleistung stimmt.*

Vorbereitung Preisverhandlung

Egal, ob Sie mit neuen Kunden verhandeln oder mit bestehenden: Bereiten Sie sich auf Preisgespräche immer vor. Stellen Sie sich folgende Fragen:

1. Wie verdient der Kunde eigentlich sein Geld? Was sind auf Sicht der nächsten Monate und Jahre Probleme und Aufgaben, auf die er zusteuert?
2. Wie hoch ist eigentlich sein tatsächliches Einkaufsvolumen? Arbeitet er beispielsweise noch mit einer weiteren Agentur zusammen?
3. Was ist der Grund, dass der Kunde den Preis verhandeln möchte – und nicht etwas anderes?
4. Was sind unsere Ziele bei der Preisverhandlung?
5. Wie lange soll der verhandelte Preis gültig bleiben?
6. Wer nimmt alles an der Preisverhandlung teil?
7. Was lief bisher gut bei der Zusammenarbeit? Was nicht?
8. Mit welchen Möglichkeiten kann die Beziehung zu den Verhandlungspartnern verbessert werden?
9. Was sind mögliche Argumente aus Kundensicht, um einen niedrigeren Preis zu bekommen?
10. Welche Sonderleistungen wurden in der Vergangenheit nicht berechnet?
11. Mit welchen Argumenten verteidigen bzw. erklären Sie Ihren Preis?
12. Welche sinnvollen Alternativen zu einem Preisnachlass können dem Kunden angeboten werden?
13. Was sind mögliche Nachteile für den Kunden, wenn er

nicht mit uns zusammenarbeitet?
14. Wie kann dem Kunden geholfen werden, Geld zu sparen oder mehr Geld zu verdienen?
15. Was ist der Plan B, wenn es bei der Verhandlungsrunde zu keinem Ergebnis kommt?

Verkäufer dürfen keine Schönredner sein

Sie sind für das verantwortlich, was Sie verkaufen, aber auch für das, was Sie nicht verkaufen. Das muss Ihnen ganz klar sein. Sie dürfen sich niemals aus der Verantwortung mit Sätzen wie: „Hätte der Kunde ja sagen können, dass er das so meint“, oder: „Hat er mir ja nicht gesagt, dass ihm x wichtig ist“, aus der Verantwortung stehlen. Denn im Zweifelsfalle bekommen Sie immer die Schuld, auch wenn Sie diese objektiv vielleicht gar nicht haben. Oder glauben Sie, dass ein Kunde, der sich von Ihnen schlecht beraten fühlt, in seinem Umfeld sagt: „Ja, und dann habe ich mich für die Agentur x entschieden. Denen habe ich leider vergessen zu sagen, dass mir a und b sehr wichtig sind. Das hat den gesamten Zeitplan zerfetzt und zu Mehrkosten von 30.000 Euro geführt“? Nein, Ihr Kunde wird in seinem Umfeld sagen: „Geh bloß nicht zu der Agentur x. Die können das nicht. Labern rum, kommen aber nicht in die Pötte. Letztlich musste ich denen noch 30.000 Euro nachschießen, weil die keine Ahnung haben.“ Klingt hart, aber so ist die Realität.
Darum müssen Sie nicht nur im Rahmen der Auftragsklärung vieles zur Verständnisförderung hinterfragen, sondern gegebenenfalls auch klare Grenzen ziehen. Beispielsweise mit Sätzen wie: „Nein, so machen wir das nicht, weil …“, oder: „Wir könnten das in so einer einfachen Version machen, wie Sie das wünschen. Das wollen wir aber nicht, weil Sie dann nicht Ihre Ziele erreichen. Und letztlich können wir es uns nicht erlauben, Sie besseren Wissens ins offene Messer laufen zu

lassen. Wir machen das also entweder richtig oder sonst lieber gar nicht. Denn wir haben auch eine Verantwortung dafür, dass Sie Ihr Geld gut investieren.“ Für manche Anbieter sind das „scharfe Aussagen“, die sie sich leider nicht zu sagen trauen. Doch wenn Sie dies wertschätzend und wohlwollend und nicht arrogant rüberbringen, dann wird so manch ein Kunde denken: „Die wissen, was sie wollen. Die scheinen ihr Geschäft zu verstehen. Also sind das wohl die richtigen Partner für mich, die ihr Geld wert sind.“

Verkaufen ist Erwartungsmanagement

Viele Kunden sind sich gar nicht darüber im Klaren, wie viel Arbeit und auch Kosten auf sie zusätzlich zukommen. Als schlechter Verkäufer werden Sie das auch von sich aus nicht kommunizieren. Als souveräner Verkäufer durchaus. Es geht nicht darum, den Kunden zu belehren, sondern ihn vielmehr zu informieren. Denn je offener die Karten sind, mit denen Sie spielen, umso harmonischer verläuft auch die Geschäftsbeziehung. Das bedeutet, dass Sie nicht nur einen klaren Zeitplan mit Ihrem Kunden gemeinsam(!) erstellen, bis wann er was geliefert haben muss, sondern dass Sie auch offensiv und rechtzeitig kommunizieren, wenn etwas von Ihrer Seite aus nicht eingehalten werden kann.

Feilschen macht Spaß

Was sagen Sie eigentlich spontan, wenn Ihr (potenzieller) Kunde zu Ihnen sagt: „Am Preis müssen wir noch was machen?“ Und Ihre Kollegen? Viele Verkäufer sind nicht in der Lage, auf diese elementare Frage souverän(!) zu antworten. So machen dann viele Anbieter ihr eigenes Unvermögen zum Problem des Kunden. Denn wenn dieser eine Antwort hört, die ihm nicht das Gefühl gibt, dass der Preis so in Ordnung ist, dann wittert er schnell die Chance auf Preisnachlässe. Es ist elementar, dass Sie nicht nur gute Antworten haben, sondern diese auch

glaubwürdig (Stichworte sind hier Stimme, Modulation, Blickkontakt, …) kommunizieren.
Es gibt viele Gründe, warum Kunden um Preisnachlässe feilschen:

- Test, ob noch „Luft“ in den Preisen ist.
- Sicherheit, nicht zu viel zu bezahlen.
- Mitbewerber bietet günstiger an.
- Einfach andere Vorstellungen vom Wert des Angebots.
- Begrenztes Budget.
- Gebot der Wirtschaftlichkeit, den besten Preis zu bekommen.
- Einkäufer haben die Aufgabe, Preise zu drücken.
- Oft kommen Nachlässe nur, weil man danach fragt.
- Spaß und Lust am Feilschen.
- Macht ausleben.

Insbesondere die letzten beiden Faktoren spielen die Hauptrolle bei Preisverhandlungen. Darum sollten Sie sich vor voreiligen Zugeständnissen hüten. Denn andernfalls denkt so manch ein Kunde: „Das war ja einfach. Ich habe einmal gefragt und bekomme jetzt schon einen Preisnachlass. Wie viel geht der denn nun noch runter, wenn ich ihn noch zwei Wochen zappeln lasse?“

> *Wenn der Kunde nicht feilscht, dann hast du zu billig angeboten.*

Antworten auf „zu teuer!“

Ein Angebot muss nicht zwangsläufig zu teuer sein, nur weil der Kunde dies sagt. Möglicherweise sind die vorgestellte Lösung und der erwartete Nutzen für den Kunden zu gering und deswegen im Verhältnis zur Leistung zu teuer. Hier ein paar Formulierungsbeispiele zur Inspiration.

1. Natürlich kann ich Ihnen beim Preis entgegenkommen. Soll ich Ihnen A oder lieber B aus dem Angebot rausrechnen?
2. Wie kommen Sie darauf?
3. Wenn Sie jetzt ein wenig mehr investieren, rechnet es sich mittelfristig besser für Sie. Schauen Sie hier …
4. Ja, billig sind wir nicht. Billig wäre es, wenn …
5. Ja, zuerst scheint der Preis ein wenig hoch zu sein. Lassen Sie uns bitte noch einmal zusammenfassen, welches Leistungspaket für Sie dahintersteckt.

Antworten auf „Mitbewerber ist billiger!“

Einkäufer müssen sich eventuell auch intern rechtfertigen, weshalb sie nicht bei dem billigsten, sondern bei einem anderen gekauft haben. Darum braucht der Einkäufer häufig nur gute Argumente vom Verkäufer, um sich intern ggf. erklären zu können. Formulieren Sie doch mal folgende Ideen in Ihren Worten um und probieren Sie aus:

1. Unser Mitbewerber ist sein Geld durchaus wert. Mehr allerdings auch nicht. Darf ich Ihnen kurz aufzeigen, was uns entscheidend von diesem abhebt?
2. Jetzt mal Hand aufs Herz. Wir würden uns doch nicht so lange unterhalten, wenn es Ihnen nur um den Preis ginge, oder?
3. Oh, das ist interessant. Haben Sie schon herausgefunden, woran die sparen, um solche Preise machen zu können? Denn auch die können ja nicht vom Verschenken leben, oder?

Über den Preis kann jeder reden

Auch wenn es respektlos klingt, aber „zu teuer!“ kann jeder

sagen. Auch jemand, der vom Thema gar keine Ahnung hat. Im schlimmsten Falle versteht also Ihr Gegenüber gar nicht, was Sie mit seinem Geld für ihn alles Gutes tun wollen. Er sieht nur Ihr Angebot, Ihren Preis – und denkt: „Das Einzige, was ich hier verstehe, ist, dass die viel Geld haben wollen. Also rede ich nur über den Preis.“
Dazu wenden manche Kunden auch interessante Tricks an. Das ist legitim. Denn auch Verkäufer nutzen ja mehr oder weniger bewusst psychologische Tricks, um den Abschluss zu machen.

- Appell über die Beziehung: Wenn man sich kennt, dann kann man doch auch nett zueinander sein. Also warum nicht einen Preisnachlass geben?
- Guter Cop, böser Cop: Einer auf der Kundenseite spielt den wohlwollenden Kunden, der andere den kritischen. Vorher haben sich beide aufgrund Ihres schriftlich eingereichten Angebots über 20.000 Euro auf 18.500 Euro verständigt. Nun fordert der böse Cop von Ihnen, vielleicht auch mit etwas Drama und spitzer Argumentation, einen Preis von 15.000 Euro. Der gute Cop vermittelt zwischen Ihnen wohlwollend und schlägt 18.500 Euro vor. Erfreut über diesen Kompromissvorschlag stimmen Sie diesem Preis zu. Ein Preiszugeständnis, welches Sie ohne den bösen Cop vermutlich niemals gegeben hätten.
- Salamitaktik: Hier fordert der Kunde immer weitere Zugeständnisse. Stimmen Sie ihm beispielsweise zu, dass die Einweisung seiner Mitarbeiter nicht berechnet wird, wird er gleich den nächsten Punkt nachverhandeln, beispielsweise die monatliche Wartungspauschale. Denken Sie, nachdem Sie hier auch entgegengekommen sind, dass es nun endlich den Auftrag gibt, wird der nächste Punkt aufs Korn genommen.

Viele Anbieter erkaufen sich Aufträge über Rabatte bzw. Leistungen, die dann „ausnahmsweise“ nicht berechnet werden.

Das rächt sich in der unternehmerischen Bilanz schnell. Denn das, was nicht eingenommen wird, kann auch nicht zum Gewinn beitragen oder für unternehmerisch wichtige Investitionen beispielsweise in die Mitarbeiterentwicklung genutzt werden.

Verkaufen Sie souverän

Alle Mitarbeiter mit Kundenkontakt müssen(!) kommunikativ geschult sein. Andernfalls besteht das Risiko, dass diese nicht nur eine Zumutung für die Kunden werden, sondern auch eine Gefahr für Ihre gesamte Unternehmung. Denn Menschen mit Kundenkontakt beeinflussen entscheidend Ihre unternehmerischen Bilanzen. Von diesen ist es abhängig, ob der Kunde den Preis leicht akzeptiert, ob er aggressiv reklamiert oder Sie mit Überzeugung weiterempfiehlt. Dass Sie Ihren Job beherrschen, setzt ein Kunde voraus. Er möchte aber die Sicherheit haben, dass Sie ein Geschäftspartner auf Augenhöhe sind und Ihren Preis wert sind.
Wer sich seiner eigenen Stärken bewusst ist, diese auch so erklären kann, dass (potenzielle) Kunden sie leicht verstehen, darüber hinaus auch wertschätzende Antworten parat hat, wenn es mal einen kritischen Einwand gibt, wird eine gute Beziehung aufbauen – und den Auftrag machen.

Datenanalysen mit R und der Google Analytics API

Datenanalysen mit R und der Google

Analytics API – websiteboosting.com

Das Programm bzw. die Programmiersprache R erfährt in den letzten Jahren eine steigende Nachfrage. Neben dem erfreulichen Aspekt, dass es Open Source und damit komplett kostenfrei ist, kann R durch Pakete flexibel erweitert werden. Neben klassischen Datamining-Paketen und fortgeschrittenen…
Das Programm bzw. die Programmiersprache R erfährt in den letzten Jahren eine steigende Nachfrage. Neben dem erfreulichen Aspekt, dass es Open Source und damit komplett kostenfrei ist, kann R durch Pakete flexibel erweitert werden. Neben klassischen Datamining-Paketen und fortgeschrittenen Analysemethoden wie neuronalen Netzen existieren diverse Möglichkeiten, den Google-Kosmos mit R anzusprechen. Die Anbindung der API von Google Analytics, Search Console bzw. der komplette Zugriff auf Sheets/Docs sind exemplarische Beispiele hierfür. Daten können in beliebiger Kombination abgefragt, aufbereitet, analysiert und visualisiert/exportiert werden.

Analysen mit R – „flexibel wie ein Schweizer Taschenmesser“

R wird seit vielen Jahren durch eine große Community kontinuierlich weiterentwickelt. Vergleichbar mit Add-ins bei Excel können in R weitere Funktionen und Schnittstellen zum Standardsystem hinzugefügt werden. So schön diese Flexibilität und Leistungsstärke klingt (und tatsächlich ist), so unspektakulär ist die Nutzerfreundlichkeit des Systems: Eingabe von Programmcode anstelle klickbarer Icons und Mausklicks à la Excel. Was jetzt als Nachteil klingt, ist gleichzeitig ein Riesenvorteil: Sofern der Programmcode erstellt ist, kann dieser jederzeit „abgespielt“ werden, die entsprechenden Analysen/Visualisierungen dauern wenige Sekunden/Minuten und können unternehmensweit geteilt werden. Jeder Schritt ist transparent und durch den Einsatz von Variablen voll flexibel – einmalig die ID des Google

Analytics-Kontos ausgeschaut und alle Analysen können automatisch von Neuem starten. R ist quasi ein „großes Excel-Makro“, das die Daten auf Befehl Schritt für Schritt verarbeitet.

Die Software R kann unter www.r-project.org/ kostenfrei heruntergeladen werden und ist nach der Installation sofort einsatzbereit. Es empfiehlt sich, zusätzlich RStudio zu installieren (www.rstudio.com/products/rstudio/download/), welches einen deutlichen Mehrwert in der Bedienung des Systems bietet (Abb. 1).

Abb. 1: RStudio „bettet“ die Software R benutzerfreundlich ein und macht die Bedienung komfortabler

Prinzipiell funktioniert R mittels Eingabe und sequenzieller Verarbeitung von Programmcode. Eingaben können sowohl in Form klassischer Rechenoperationen erfolgen (bspw. 3+3) als auch als Definition von Formeln, Funktionen, Variablen, Vektoren und Dataframes. Der Vorteil der Nutzung von Variablen ist, dass diese nach Definition zur gesamten Laufzeit zur Verfügung stehen und im weiteren Verlauf darauf zurückgegriffen werden kann. Füllt man bspw. die Variable „ga_id“ (Variablennamen sind frei wählbar) mit der Datenansichts-ID von Google Analytics und verweist im weiteren Verlauf der Berechnungen auf die Variable anstelle einer fix definierten Nummer, kann die identische Analyse durch Neudefinition der Variable flexibilisiert werden. Variablen werden in R mittels der Syntax „Variablenname <- Variableninhalt“ definiert.

googleAnalyticsR – das flexible R-Paket

Damit R auf die API von Google Analytics zugreifen kann, muss ein Paket installiert werden (genereller Befehl: install.packages(„Paketname“). Die Installation des Pakets erfolgt nach Befehlseingabe binnen Sekunden, da die Pakete

automatisch aus dem Internet geladen werden. Die Erweiterung „dependencies = TRUE“ bewirkt, dass passende Pakete zusätzlich installiert werden, insb. googleAuthR zur OAuth-Authentifizierung des Google-Nutzerkontos mit R (siehe Abb. 2).

Abb. 2: Einfache Installation von googleAnalyticsR nebst weiteren Paketen direkt in R

Nachdem die gewünschten Pakete installiert wurden, stehen diese prinzipiell zur Verfügung. Damit sie tatsächlich auch genutzt werden können, müssen sie mit dem Befehl library(„Paketname“) geladen werden. Das heißt, nach Absetzen des Befehls „library(“googleAnalyticsR“)“ besteht eine Verbindung von R zur Google Analytics API. Der Befehl „ga_auth()“ bzw. „gar_auth()“ aus dem Paket googleAuthR ermöglicht die Authentifizierung von R mit dem jeweiligen Google-Konto über das OAuth-Verfahren. Ein erstmaliger Aufruf leitet zum Google-Konto-Log-in nebst Angabe der übergebenen Berechtigungen (siehe Abb. 3).

Abb. 3: Zugriff des R-Pakets auf Google-Konto

Sofern das Google-Konto bei Google Analytics mindestens das Recht „Bearbeiten“ auf eine Property hat, kann im weiteren Verlauf direkt die Google API genutzt werden. Wie angesprochen ist es sinnvoll, mit Variablen zu arbeiten und die gewünschte Auswahl an Metriken und Dimensionen in Vektoren zu speichern. Der Befehl „ga_account_list()“ erzeugt eine Übersicht aller Analytics Propertys nebst Datenansichten (Views) im jeweiligen Google-Konto. Als Beispiel kann eine Variable „account_list“ mit genau dieser Übersicht befüllt (Befehl account_list <- ga_account_list()) und anschließend aufgerufen werden (siehe Abb. 4). Auf Basis dieser Liste kann bspw. eine Variable „ga_id“ dynamisch mit der zweiten Datenansicht befüllt werden. Hierzu wird das Feld viewId mit dem Dollarzeichen ($) aus account_list angesprochen und die zweite Position adressiert. Alternativ könnte die Variable ga_id selbstverständlich direkt

mit der jeweiligen View-Nummer befüllt werden.

Abb. 4: R-Befehle zum Abruf von R

Mit dem Befehl „google_analytics“ wird die API angesprochen und die entsprechenden Werte werden in R ausgegeben. Sofern bspw. der Datumbereich sowie die gewünschten Metriken und Dimensionen nebst der View-ID im Befehl google_analytics übergeben werden, werden die Daten unmittelbar abgerufen. Aufgrund des Abrufs der Daten über die API kann ggf. das Sampling von Daten umgangen werden. Hierzu bedarf es beim Aufruf des Befehls google_analytics der Ergänzung um den Parameter anti_sample = TRUE. Die Funktion teilt dabei den API-Aufruf in mehrere Teilaufrufe auf, damit möglichst ungefilterte Daten über die API abgerufen werden. Weiterhin kann mit dem Parameter max = „Zahl“ die maximale Anzahl von abzurufenden Daten limitiert werden (siehe Abb. 5).

Tipp: Es besteht für Google Analytics ein sog. „Dimensionen- und Metriken-Explorer“, der anzeigt, welche Dimensionen und Metriken generell kombiniert werden dürfen und wie die entsprechenden Felder in der API heißen (bspw. ga:bounceRate für die Absprungrate in %): einfach.st/gareferenzen.

Das Paket googleAnalyticsR nutzt ein allgemeines Kontingent an API-Aufrufen von Google Analytics. Bei intensiver Nutzung und zur Sicherstellung der Verfügbarkeit sollte ein eigenes API-Kontingent über Google Developers beantragt werden (https://console.developers.google.com). Nach Registrierung bei Developers stehen diverse Google APIs zur Verfügung. Als Anpassung in R müssen vor der Authentifizierung die Client-ID sowie der geheime Clientschlüssel als Option angegeben werden (options(googleAuthR.client_id = „xx.apps.googleusercontent.com“) sowie options(googleAuthR.client_secret = „xx“). Beide Informationen sind nach der Registrierung und Aktivierung der Analytics API abrufbar. Bei Google Analytics bietet die API 50.000 Abfragen pro Tag/User bzw. alle 100 Sekunden 100 Abfragen.

Flexibler Export und Datenvisualisierungen mit ggplot2

Sofern das Ergebnis des Abrufs mittels „google_analytics“ in eine Variable/Dataframe gespeichert wird, können neben der Durchführung statistischer Auswertungen (bspw. summary(„Variable“) die Daten in ein spezifisches Dateiformat gespeichert und dann mit anderen Programmen weiterverarbeitet werden. Eine CSV-Datei wird über den Befehl „write.csv“ erzeugt (siehe Abb. 5).

Tipp: Perfekt harmoniert R auch mit Google Sheets, was damit mittelbar eine nachgelagerte Datenvisualisierung mit Google Data Studio ermöglicht. Das Szenario wäre folgendes: Daten werden aus Google Analytics und anderen Quellen in R analysiert und verarbeitet, automatisch über das R-Paket „googlesheets“ in ein Google Spreadsheet übertragen und anschließend über Data Studio visualisiert und ggf. mit weiteren Datenquellen ergänzt.

Abb. 5: Daten aus der API in R analysieren und als .csv speichern

R hat herausragende Möglichkeiten der Datenvisualisierung. Es existiert eine Vielzahl an Paketen, welche die Daten mit weiteren Analysen grafisch anreichern. Pakete wie „RColorBrewer“ erzeugen u. a. Heatmaps als Datenvisualisierung. Das Paket „ggplot2“ (install.packages(“ggplot2“) sowie Aktivierung über library(“ggplot2“)) bietet Schaubilder, die in Excel nicht existent sind. Mittels des Befehls geom_boxplot wird bspw. ein Boxplot erstellt, geom_line / geom_smooth erzeugt ein Liniendiagramm nebst nicht linearer Trendlinie (siehe Abb. 6).

Abb. 6: Datenvisualisierungen mit dem Paket ggplot2

Tipp: Die Visualisierung bzw. Berichterstellung ist in R sehr

umfassend durch sogenanntes R Markdown nutzbar. Es wird einmalig der Programmcode erstellt bzw. ein Bericht konzipiert, welcher anschließend als HTML-Seite, PDF oder Word-Dokument ausgegeben wird. Damit sind aufwendiges „Exceln“ und formatieren obsolet, die Corporate Identity ist in hohem Maße in den Berichten standardisierbar.

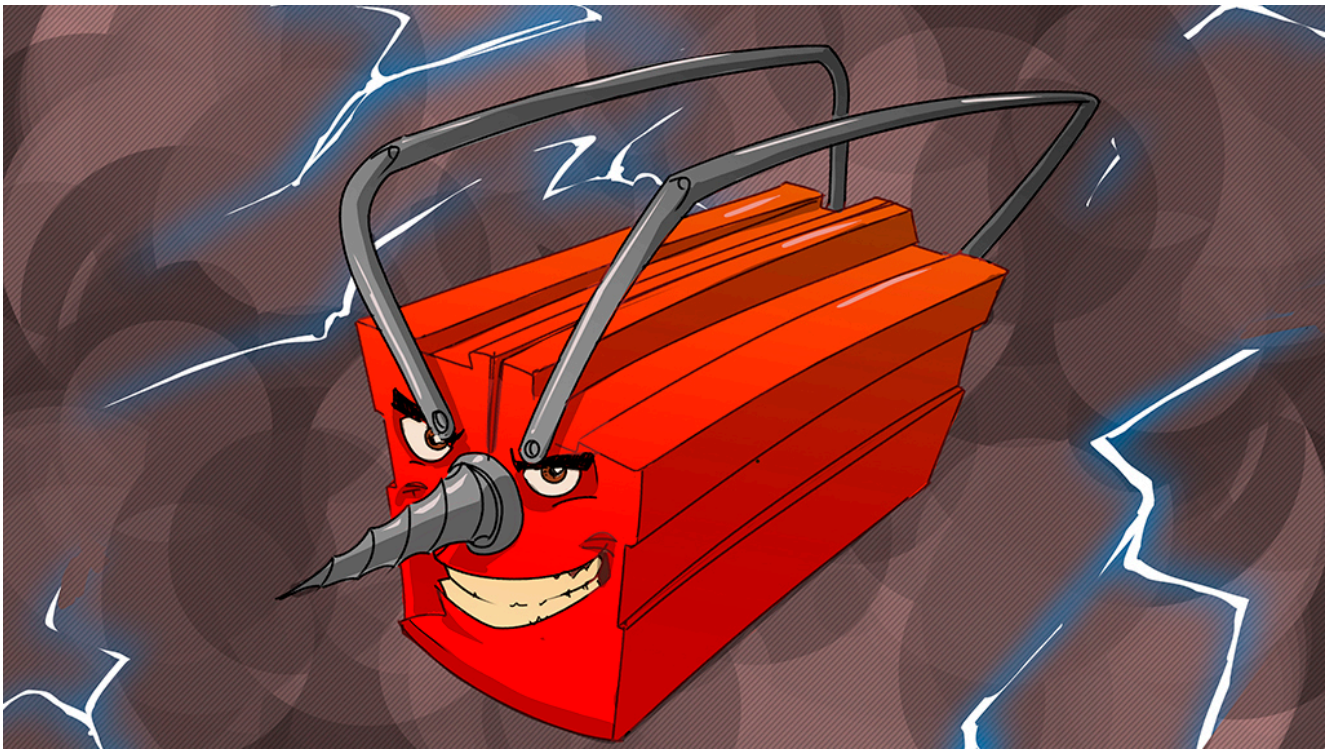
Richtig interessant werden Analysen und Visualisierungen, wenn diese auf Rohdaten beruhen. Mittels der Analytics API können viele Informationen verarbeitet werden, die es später zu clustern gilt. Eine sinnvolle Ergänzung ist deshalb, die Google-Analytics-Daten durch benutzerdefinierte Dimensionen zu erweitern. Wird jedem Datensatz ein exakter Zeitstempel mitgegeben, jede Session mit einer Session-ID in den Daten ergänzt sowie die Cookie-ID als Klammer für den Besucher verstanden, so könnte eine Customer-Journey-Analyse mit Rohdaten erfolgen. Simo Ahava hat in seinem Blog eine detaillierte Anleitung für diese benutzerdefinierten Dimensionen erstellt (http://einfach.st/simohava). Es ist sehr ratsam, vor der Integration den jeweiligen Datenschutzverantwortlichen zu konsultieren!

In diesem Sinne: Geben Sie R eine Chance! Im ersten Schritt hat es eine „Old-School“-Anmutung, da Befehle über eine Console eingegeben werden müssen. Mit Neugier und Training werden die Abläufe Tag für Tag vereinfacht und Webanalyse bzw. Datenvisualisierung kann wirklich auf „Knopfdruck“ erfolgen – so besteht viel mehr Freiraum und Zeit, über die Ergebnisse nachzudenken bzw. diese zu interpretieren, anstatt Zeit für die Datenaufbereitung ineffizient zu vergeuden. Und zuallerletzt: Das System wird permanent (kostenlos) weiterentwickelt– für jede Anforderung existiert bestimmt ein Paket.

Noch ein Literaturhinweis: Das kostenlose englischsprachige E-Book „Using Google Analytics with R“ von Michal Brys ist unter http://michalbrys.com/book/ in diversen Formaten abrufbar und bietet einen hervorragenden Einstieg in die Nutzung der

Analytics API!

Rechtliche Unsicherheiten bei Hacking-Werkzeug



Kommt drauf an, wozu

Um Schwachstellen im eigenen Netz zu suchen und Datenflüsse zu überwachen, nutzen Administratoren vielfach die gleichen Softwarehilfsmittel wie Angreifer, die illegale Ziele verfolgen. Was sagt das deutsche Recht dazu? Steht jemand, der mit trickreichen Analyse- und Spähtools arbeitet, mit einem Bei…

Kommt drauf an, wozu

Rechtliche Unsicherheiten bei Hacking-Werkzeug

Um Schwachstellen im eigenen Netz zu suchen und Datenflüsse zu überwachen, nutzen Administratoren vielfach die gleichen Softwarehilfsmittel wie Angreifer, die illegale Ziele verfolgen. Was sagt das deutsche Recht dazu? Steht jemand, der mit trickreichen Analyse- und Spähtools arbeitet, mit einem Bein vor Gericht?

Von Verena Ehrl

Der Theaterautor und Sprachliebhaber Hans-Joachim Haecker brachte die Dual-Use-Problematik bereits 1968 in einem Gedichtchen seines Bandes „Insonderheit“ unter dem Eindruck des internationalen Wettrüstens scherzhaft auf den Punkt:

„Insonderheit die Abwehrwaffen
sind für die Abwehr wie geschaffen.
Auch kann man mit geschickten Händen
sie für den Angriff gut verwenden.“

Die Waffen, mit denen Akteure innerhalb der IT-Welt hantieren, eignen sich zum Eindringen in Systeme, zum Überwinden von Sicherungsmaßnahmen, zum Spionieren und Manipulieren. Dieselben Werkzeuge können aber dazu dienen, unterschiedliche Ziele zu erreichen. In der Hand eines Administrators, der ein System, für das er verantwortlich ist, Penetration Tests („Pentests“) aussetzt, kann etwa das Software-Tool „Mimikatz“ legalen Einsatz finden (S. 24). Es ebnet allerdings ebenso gut Angreifern den Weg bei illegalen Aktionen, indem es ihnen Zugangsdaten für die Übernahme eines Netzwerks offenbart. Auf diese Ambivalenz bezieht sich das Schlagwort „Dual Use“ im Zusammenhang mit Hackerausrüstung.

Nicht immer erscheinen die Werkzeuge, um die es geht, so spektakulär wie die Hacking-Gadget-Hardware, mit der c’t sich

in Ausgabe 18/2017 befasst hat [1]. Sehr oft steht vielmehr bloße Software im Mittelpunkt, die loggt, sucht, entschlüsselt, ausliest, analysiert und speichert (S. 16, 18, 24 und 39). Die rechtlichen Fragen, vor die ein Anwender gestellt ist, sind jedoch grundsätzlich die gleichen wie bei den Spionagegerätchen [2]: Wo liegt die rote Linie, jenseits der man sich auf illegalem Terrain bewegt? Was sagt das geltende Recht zum Umgang mit potenziell gefährlichen und schadenträchtigen Tools?

Dass es „Güter mit doppeltem Verwendungszweck“ gibt, die sich gleichermaßen für legales und illegales Tun eignen, beschäftigt nicht zuletzt den europäischen Gesetzgeber. Am 9. September 2021 ist die Neufassung der sogenannten Dual-Use-Verordnung in Kraft getreten [3].

Sie betrifft „Güter einschließlich Datenverarbeitungsprogramme (Software) und Technologie, die sowohl für zivile als auch für militärische Zwecke verwendet werden können“. Es gibt durchaus Softwareentwicklungsprojekte, die man mit etwas Fantasie in den Betrachtungshorizont der Verordnung rücken kann.

Ziel der Dual-Use-Verordnung ist die Exportkontrolle. Die kann bereits relevant werden, wenn Forschung und Produktentwicklung mit Partnerunternehmen in bestimmten außereuropäischen Ländern stattfinden und dabei schadenträchtige Software im Spiel ist.

```
mimikatz 2.2.0 x64 (oe.eo)
  .#####.   mimikatz 2.2.0 (x64) #19041 Aug 10 2021 17:19:53
 .## ^ ##.  "A La Vie, A L'Amour" - (oe.eo)
 ## / \ ##  /*** Benjamin DELPY `gentilkiwi` ( benjamin@gentilkiwi.com )
 ## \ / ##       > https://blog.gentilkiwi.com/mimikatz
 '## v ##'       Vincent LE TOUX             ( vincent.letoux@gmail.com )
  '#####'        > https://pingcastle.com / https://mysmartlogon.com ***/

mimikatz # privilege::debug
Privilege '20' OK

mimikatz # log
Using 'mimikatz.log' for logfile : OK

mimikatz # sekurlsa::logonpasswords

Authentication Id : 0 ; 256573 (00000000:0003ea3d)
Session           : Interactive from 1
User Name         : rei
Domain            : Asus11
Logon Server      : ASUS11
Logon Time        : 11.10.2021 12:07:31
```

Mimikatz ist ein typisches Beispiel für ein Hackerwerkzeug, das auch verantwortungsvollen Admins wertvolle Dienste leistet, wenn es ums Aufspüren von Schwachstellen im eigenen Netz geht.

Neu im Blick der europäischen Rechtssetzungsorgane und der zur Umsetzung verpflichteten Mitgliedsstaaten sind Länder, welche die Todesstrafe praktizieren oder in denen Menschenrechtsverletzungen wie Folter auf staatliches Geheiß stattfinden. Wer etwa potenziell gefährliche Software ins nichteuropäische Ausland transferieren will, braucht dazu eine vorherige Genehmigung des Bundesamts für Wirtschaft und Ausfuhrkontrolle (BAFA). Diese Behörde entscheidet in jedem Einzelfall, ob der Transfer zulässig ist oder nicht.

Zweierlei Paar Schuhe

Abseits der von der Verordnung erfassten Transferproblematik sind Dual-Use-Softwarewerkzeuge rechtlich schwer zu fassen. Weder ihr Erwerb noch ihr Besitz ist grundsätzlich untersagt. Straf- und Zivilrecht melden sich erst dann, wenn jemand diese Tools einsetzt, um Rechtsbrüche zu begehen. Derjenige riskiert dann eine Strafe oder er sieht sich zivilrechtlichen Ansprüchen ausgesetzt – oft droht ihm beides.

Nicht alles, was jemand rechtswidrig tut, ist auch strafbar. Mit dem Strafrecht geht der Staat gegen von ihm untersagtes Verhalten vor, dabei sind Strafermittlungsbehörden im Spiel, es gibt Beschuldigung und Anklage. Im Zivilrecht stehen hingegen gleichberechtigte Streitparteien einander gegenüber. Dabei gibt es Kläger und Beklagte, die Gerichte entscheiden über Ansprüche, welche die Parteien gegeneinander geltend machen. Vertragspflichten und Schadenersatzansprüche sind typisches zivilrechtliches Geschäft.

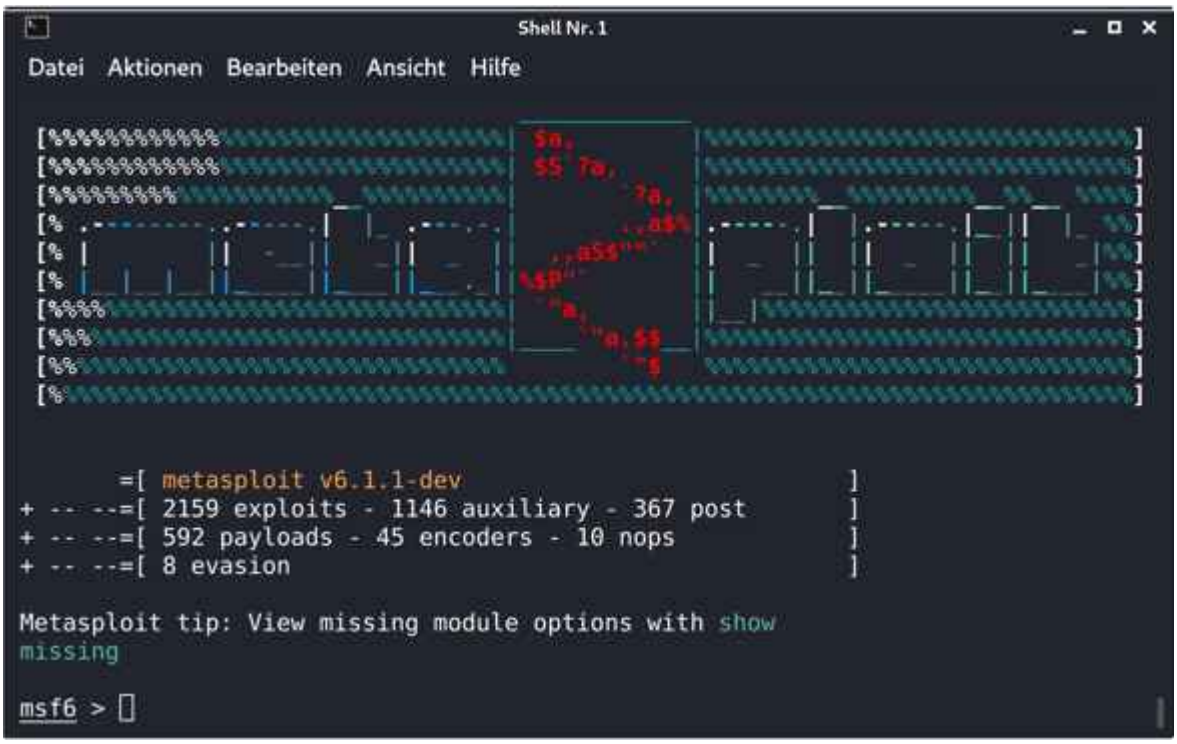
Durch Software kann ein Anwender sich Ärger in beiden Rechtsbereichen einhandeln. Ein gutes Beispiel für die rechtliche Gratwanderung dabei sind die bereits angesprochenen Pentests. Sie haben die Aufgabe, Schwachstellen in Konfiguration, Hard- und Software von Servern, Routern und Arbeitsplätzen im Netz aufzuzeigen. Ein Mitarbeiter, der einen solchen Test im Auftrag eines zuständigen Entscheiders für das betroffene Netz durchführt, bewegt sich damit auf der legalen Seite. Wenn allerdings etwa ein Cybersecurity-Dienstleister einen Pentest unaufgefordert und ohne Erlaubnis bei einem potenziellen Kunden durchführt, um diesen als Auftraggeber zu gewinnen, setzt er sich strafrechtlicher Verfolgung wegen Datenveränderung oder Computersabotage aus.

Wesentlich sind dabei die Paragrafen 303a und 303b des Strafgesetzbuchs (StGB), die sich mit virtueller Sachbeschädigung befassen. Damit hat der Gesetzgeber 1986 eine Regelungslücke beim Straftatbestand der Sachbeschädigung (§ 303 StGB) geschlossen. Die klassische Sachbeschädigung setzt einen körperlichen Gegenstand voraus – das lässt Daten und beispielsweise Festplatten, die funktionsfähig bleiben, aber deren Inhalt gelöscht oder verschlüsselt wurde, außen vor.

§ 303a StGB stellt die unbefugte Veränderung von Daten unter Strafe. Dort heißt es in Absatz 1: „Wer rechtswidrig Daten (§ 202a Abs. 2) löscht, unterdrückt, unbrauchbar macht oder verändert, wird mit Freiheitsstrafe bis zu zwei Jahren oder mit Geldstrafe bestraft.“

Bereits der Versuch ist strafbar; dasselbe gilt wie beim Ausspähen (§ 202a StGB) und Abfangen von Daten (§ 202b StGB) auch fürs „Vorbereiten“ einer rechtswidrigen Datenveränderung – beispielsweise durch Bereitstellen von Software, deren „Zweck die Begehung einer solchen Tat ist“ (§ 202c Abs. 1 Nr. 2 StGB).

Als Daten in diesem Sinne gelten nach § 202 StGB „nur solche, die elektronisch, magnetisch oder sonst nicht unmittelbar wahrnehmbar gespeichert sind oder übermittelt werden“, also nicht etwa Papiernotizen.



Das modulare Pentesting-Tool Metasploit ist bei Security-Experten beliebt, es lässt sich aber auch für echte Angriffe missbrauchen.

§ 303b StGB, der sogenannte Hackerparagraf, ist der Computersabotage gewidmet. Er steht in direktem Zusammenhang mit seinem Nachbarn, der die Datenveränderung betrifft, und lautet: „Wer eine Datenverarbeitung, die für einen anderen von wesentlicher Bedeutung ist, … erheblich stört, … wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.“

Wenn es um die Datenverarbeitung eines Unternehmens respektive

einer Behörde geht, stehen Freiheitsstrafen bis zu fünf Jahren im Raum. In besonders schweren Fällen riskieren Täter sogar bis zu zehn Jahren Gefängnis. Der Paragraf grenzt die Wege, auf denen die strafbare Datenverabeitungsstörung bewirkt wird, näher ein: Unter Nr. 1 fasst er die Störung durch Löschen, Unterdrücken, Unbrauchbarmachen oder Verändern von Daten (siehe § 303a Abs. 1). Unter Nr. 2 erscheint die Dateneingabe und -übermittlung mit der Absicht, jemand anderem einen Nachteil zuzufügen. Unter Nr. 3 schließlich geht es darum, dass eine Datenverarbeitungsanlage oder ein Datenträger zerstört, beschädigt, unbrauchbar gemacht, beseitigt oder verändert wird. Auch hierbei sind wie bei der Datenveränderung bereits Versuch und Vorbereitung strafbar.

Dass es nur ums Stören von Datenverarbeitungen von „wesentlicher Bedeutung“ geht, soll Bagatellfälle aus dem Blick der Strafjustiz nehmen.

Erpressungstrojaner vor Gericht

Im April 2021 hat der Bundesgerichtshof (BGH) einen Fall entschieden, der die Verteilung von Ransomware betraf – also das Einschleusen von Verschlüsselungstrojanern zu Erpressungszwecken [4]. Dabei stufte das Gericht unter anderem das Anbringen einer Eintragung in der Windows-Registry, die das automatische Laden einer Schadsoftware beim Rechnerstart bewirkte, als strafbare Datenveränderung ein. Zugleich sah der BGH in diesem Fall auch eine Computersabotage nach § 303b StGB.

Der Angeklagte war Mitglied einer Cybercrime-Bande mit Sitz in der Ukraine. Diese hatte mit ihren Ransomware-Angriffen von 2013 bis 2016 über 200 Millionen Rechnersysteme infiziert und von den geschädigten Computernutzern mehr als neun Millionen Euro erpressen können.

Schadsoftware ähnlicher Art mit einer vorgesehenen Entsperrmöglichkeit hätte aber auch im Rahmen eines Pentests

in einem lokalen Netz durchaus legal verwendet werden können. Ein Unternehmen hätte damit etwa seine Mitarbeiter auf deren Vorsicht testen können, was das Anklicken unbekannter Inhalte betrifft.

Computersabotage am Arbeitsplatz

Das mutwillige Stören des Datenverarbeitungsbetriebs kann nicht nur strafrechtliche, sondern auch arbeitsrechtliche Auswirkungen haben. Mitarbeiter, die aus Wut oder infolge von Rachegedanken gegen ihren Arbeitgeber Datenvandalismus im Unternehmensnetz betreiben, riskieren eine fristlose Kündigung.

Das geschah 2019 einem als Key-Account-Manager beschäftigten Arbeitnehmer nach einer heimlichen Löschaktion auf dem Unternehmensserver seines Arbeitgebers. Nach einer Abmahnung war ihm zuvor ein Aufhebungsvertrag angeboten worden. Daraufhin löschte er 8 GByte an Daten, darunter Kalkulationssoftware, Umsatzmeldungen, Vorlagen für Preislisten und Wettbewerbsanalysen für bestimmte Produkte. Die Daten konnten später wiederhergestellt werden. Der Verdacht fiel auf ihn. Gegen die fristlose Kündigung , die er wenig später erhielt, wehrte er sich zunächst erfolgreich. In der Berufungsinstanz jedoch unterlag er vor dem Landesarbeitsgericht (LAG) Baden-Württemberg im September 2020 [5]. Sein Arbeitgeber hatte einen 93seitigen Vergleich des Datenbestands im fraglichen Verzeichnis vor und nach der Löschaktion vorgelegt.

Das Gericht sah die fristlose außerordentliche Kündigung als begründet an: Das unbefugte vorsätzliche Löschen betrieblicher Daten auf EDV-Anlagen des Arbeitgebers tauge ebenso wie das Vernichten von Verwaltungsvorgängen grundsätzlich als „wichtiger Grund“ für eine solche Kündigung im Sinne des § 626 Abs. 1 BGB. Dabei komme es nicht unbedingt darauf an, ob sich der Arbeitnehmer nach § 303a StGB oder § 303b StGB strafbar gemacht habe. Es sei auch nicht entscheidend, ob und mit

welchem Aufwand ein Teil der gelöschten Daten wiederhergestellt werden konnte oder ob der Arbeitgeber diese Daten für den weiteren Geschäftsablauf tatsächlich benötigte. Vielmehr gehöre es zu den vertraglichen Nebenpflichten eines Arbeitsverhältnisses im Sinne des § 241 Abs. 2 BGB, dass der Arbeitnehmer seinem Arbeitgeber den Zugriff auf betriebliche Dateien nicht verwehre oder unmöglich mache.

In der Hand von Angreifern kann auch der Fern Wifi Cracker, mit dem man Drahtlosnetze auf Sicherheitslücken abklopft, zum Werkzeug für eine Straftat werden.

Wenn der Admin spioniert

Dass es bei der strafrechtlichen Bewertung von Hacker-Aktivitäten nicht so sehr um die benutzten Werkzeuge als vielmehr um Zweck und Absicht des Einsatzes geht, illustriert auch ein Fall, den 2020 der BGH in letzter Instanz entschied [5]. Zwei Männer mussten sich wegen des Ausspähens von Daten (§ 202a StGB) verantworten. Einer davon leitete die Stabsstelle eines Apothekerverbandes und betrieb daneben ein gesundheitspolitisches Informationsportal im Internet. Der zweite Angeklagte arbeitete als Systemadministrator am Berliner Standort des Bundesgesundheitsministeriums (BMG) und war nebenbei als Callboy tätig – daher rührte auch die Bekanntschaft der beiden Männer.

Der Admin hatte jahrelang Zugriffsrecht auf alle E-Mail-Accounts seiner Dienststelle und versorgte den Portalbetreiber mit so gewonnenen Interna aus dem Ministerium. Nachdem das Ministerium den unbeschränkten Mailzugriff der Administratoren im Hause als Sicherheitsmangel erkannt hatte, wurde es für den Mann schwieriger – er verlegte sich schließlich auf einen unter den Admins bekannten Notfalltrick, mit dem er sich selbst von Fall zu Fall die nötigen Zugriffsrechte verschaffte. Er lieferte dem Portalbetreiber auf Datenträger kopierte E-Mail-Inhalte nach Wunsch und kassierte dafür insgesamt rund 1000 Euro. Sein Kunde war besonders an E-Mails der Minister und Staatssekretäre sowie einiger Abteilungs- und Referatsleiter interessiert. In den weitergereichten Mails ging es unter anderem um Gesetzesvorhaben, die für das Publikum des Portals besonders interessant waren.

Das Landgericht (LG) Berlin verurteilte die Männer im April 2019 wegen gemeinschaftlichen Ausspähens von Daten nach § 202a StGB und sah in der Manipulation der Zugriffsrechte auf die einzelnen E-Mail-Konten zudem die Überwindung einer Zugangssicherung nach § 202a Abs. 1 StGB. Die gegen das landgerichtliche Urteil eingelegte Revision wies der BGH

weitgehend ab, lediglich den einzuziehenden Geldbetrag setzte er niedriger an als die Vorinstanz.

Die Bundesrichter beschäftigten sich vor allem damit, ob Daten im Sinn des §202a Abs. 1 StGB überhaupt als gesichert gelten können, wenn ein Admin mithilfe seiner Kenntnisse darauf zugreifen kann. Die Antwort: Es genüge, wenn getroffene Vorkehrungen den Zugriff auf Daten zumindest deutlich erschweren. Die Sicherung von E-Mail-Accounts durch Passwörter reiche aus. Dabei brauche der Systembetreiber nur den Zugriff Unbefugter zu berücksichtigen, aber nicht die Zugriffsmöglichkeit durch Eingeweihte oder Experten. Es sei nicht erforderlich, dass die Sicherung gerade gegenüber dem Täter wirkt – wenn dieser etwa ein Administrator ist, der den tatsächlichen Zugriff auf die Daten hat.

Als Überwinden der Zugangssicherung nach § 202a StGB können dem Gericht zufolge auch Handlungen gelten, die nicht besonders anspruchsvoll oder aufwendig sind. Wenn jemand durch Insiderkenntnisse oder Ähnliches eine Absicherung schnell und leicht ausschalten kann, zähle das rechtlich ebenso, als hätte sich jemand durch raffinierte technische Werkzeuge Zugriff verschafft. Nur wenn eine Durchbrechung des Schutzes für jedermann ohne Weiteres möglich sei, werde der Tatbestand nicht erfüllt.

Wer den Schaden hat …

Wie bereits gesagt, ist strafrechtliche Verfolgung nicht das Einzige, was der illegale Einsatz von Hacking-Tools nach sich ziehen kann: Wenn dabei ein Schaden entsteht, hat der Geschädigte einen Anspruch auf Schadenersatz gegen den Verursacher. Schäden durch IT-Störmanöver können sehr hoch sein – wenn etwa durch den Ausfall von Unternehmensservern Arbeitsprozesse lahmgelegt werden. Auch der Ausfall der Netzkommunikation kann enorme Umsatzeinbußen und damit hohe wirtschaftliche Schäden bedeuten. Für Anspruchsteller im Zivilrecht ist wichtig, dass jede Streitpartei alles, was für

ihre Sache spricht, selbst gerichtsfest beweisen muss. Das kann bei Schäden durch Hackertools etwa den Nachweis eines Hackerangriffs und die zweifelsfreie Benennung des Angreifers betreffen. Außerdem ist auch nachzuweisen, dass der Angriff tatsächlich den geltend gemachten Schaden hervorgerufen hat.

Wenn ein Täter bereits strafrechtlich wegen einer Computerstraftat zulasten eines Geschädigten verurteilt worden ist, hat jener es anschließend vergleichsweise leicht, seine Ansprüche gegen den Verurteilten zivilrechtlich geltend zu machen: Das Urteil des Strafgerichts hat selbst bereits Indizwirkung, zudem kann der Kläger die im Strafverfahren erhobenen Beweise zu seinen Gunsten nutzen.

Ein dritter, bislang noch nicht genannter Bereich, der durch unrechtmäßigen Einsatz von Hackertools berührt werden kann, ist der Datenschutz. Wo bei einem Angriff personenbezogene Daten im Spiel sind, geht es nicht bloß um wirtschaftliche Schäden, sondern möglicherweise auch um die massenhafte Verletzung des informationellen Selbstbestimmungsrechts der Betroffenen. Wer es also als Pentester mit realen Datenbeständen zu tun hat, tut gut daran, die Bestimmungen der europäischen Datenschutzgrundverordnung (DSGVO) sorgfältig zu beachten. *(psz@ct.de)*

1. Literatur

2. Ronald Eikenberg, David Wischnjak, Böse und billig: Hacking-Gadgets, Gefahr durch angriffslustige Hardware, c’t 18/2017, S. 62 und S. 64
3. Verena Ehrl, Elektronische Übeltäter, Rechtliche Aspekte im Zusammenhang mit Spionage- und Sabotage-Gadgets, c’t 18/2017, S. 78
4. Verordnung (EU) 2021/821 des Europäischen Parlaments und des Rates vom 20.5.2021 über eine Unionsregelung für die Kontrolle der Ausfuhr, der Vermittlung, der technischen Unterstützung, der Durchfuhr und der Verbringung betreffend Güter mit doppeltem Verwendungszweck:

heise.de/s/N76o
5. BGH, Beschluss vom 8.4.2021, Az. 1 StR 78/21: heise.de/s/rmDJ
6. LAG Baden-Württemberg, Urteil vom 17.9.2020, Az. 17 Sa 8/20: heise.de/s/MXBL
7. BGH, Beschluss vom 13.5.2020, Az. 5 StR 614/19: heise.de/s/Zvpw

---

WordPress – Was das Content-Management-System so einzigartig macht

Loslegen mit WordPress

Egal, ob Sie der Welt Bilder Ihrer Katzen präsentieren, Ihrem Unternehmen eine virtuelle Heimstatt bauen oder sogar online

Produkte verkaufen wollen – WordPress kanns. Das Content-Management-System ist vielfältig anpassbar und eine riesige Nutzer- und Entwicklergemeinde steht dahinter. Diese Artikel… Egal, ob Sie der Welt Bilder Ihrer Katzen präsentieren, Ihrem Unternehmen eine virtuelle Heimstatt bauen oder sogar online Produkte verkaufen wollen – WordPress kanns. Das Content-Management-System ist vielfältig anpassbar und eine riesige Nutzer- und Entwicklergemeinde steht dahinter. Diese Artikelstrecke macht den Einstieg zum Kinderspiel

Von Jo Bager

Ein Content-Management-System (CMS) dient dazu, die Inhalte, das Layout und die Funktionen einer Website zu verwalten. Das mit Abstand beliebteste CMS ist WordPress. Mehr als 40 Prozent der Websites, die die Web-Vermesser von w3techs.com analysieren, setzen es ein. Das System ist zwar (immer noch) das Mittel der Wahl vieler Blogs, aber längst nicht mehr nur in dieser Nische zu Hause: WordPress steht hinter so riesigen Websites wie dem Tech-Magazin Techcrunch (techcrunch.com), Microsoft News (news.microsoft.com) oder der Homepage des Rolling Stone Magazine (rollingstone.com).

2004 hatte der Web-Entwickler Matt Mullenweg die erste stabile Version von WordPress herausgebracht, ursprünglich als dediziertes Blog-System. Dass das CMS heute so dominant ist, hat mehrere Gründe. So hat Mullenweg erstens schon 2005, also relativ früh, einen kostenpflichtigen WordPress-Hosting-Dienst gestartet (wordpress.com), auf dem jedermann auch heute noch WordPress-Präsenzen aufsetzen kann. Das ermöglicht auch technischen Laien einen einfachen Einstieg und der direkte Kontakt zu Kunden war sicherlich ein guter Wegweiser für die Weiterentwicklung von WordPress.

Die Software selbst blieb und bleibt – zweitens – kostenlos und als Open Source verfügbar. Dafür steht eine US-amerikanische gemeinnützige Stiftung ein, der Mullenweg vorsteht. Diese WordPress Foundation betreibt unter anderem

wordpress.org (nicht .com), das Zuhause der Software. Drittens wurde WordPress konsequent weiterentwickelt. Zu den Blog-Funktionen kam etwa die Unterstützung statischer Seiten hinzu, die klassische Websites erst möglich machte.

Bei den Funktionen wurde – viertens – schon sehr früh auf Benutzerfreundlichkeit geachtet, wodurch WordPress zum Beispiel recht einfach zu installieren ist. Vor allem aber war der Funktionsumfang von WordPress – fünftens – schon seit ziemlich frühen Versionen durch Plug-ins erweiterbar und auch das Aussehen lässt sich bereits seit Langem mit Themes anpassen. So ist WordPress nach und nach zu einer Plattform gewachsen, mit Abertausenden Themes und fast 60.000 Plug-ins.

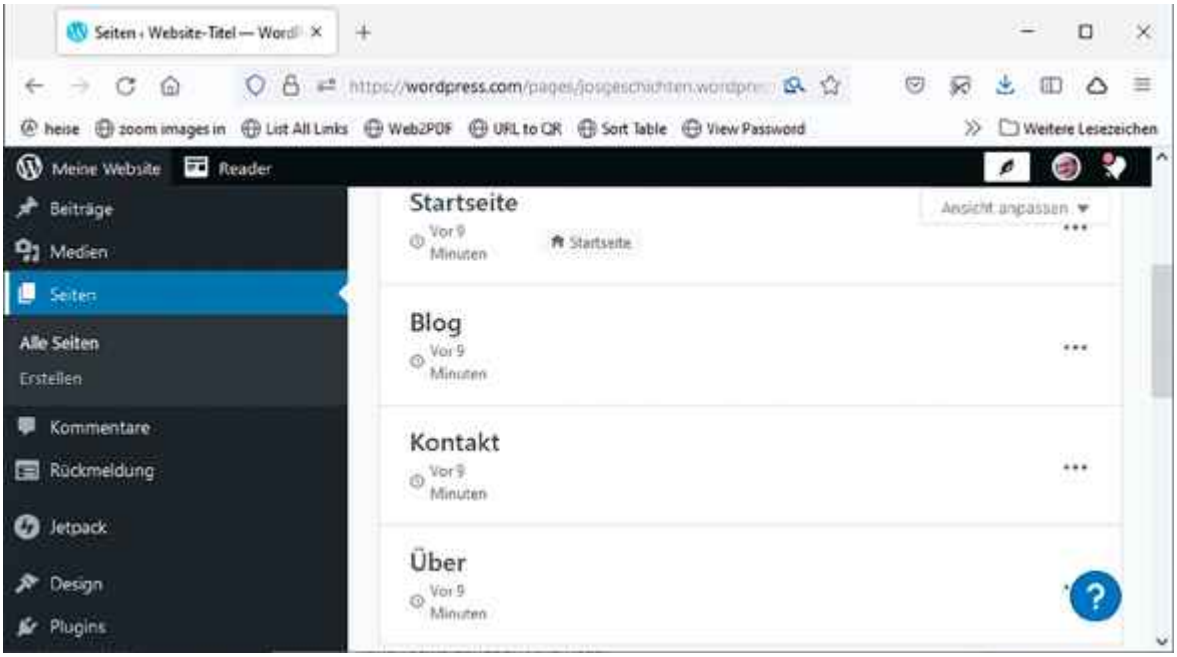
Ein gewaltiges Ökosystem

Rund um WordPress ist eine riesige Biosphäre entstanden: Sie finden in den Plug-in- und Theme-Verzeichnissen nicht die gewünschte Funktion, den passenden Look? Sie haben ein technisches Problem, Ihre WordPress-Site ist zum Beispiel zu langsam? Kein Problem, es finden sich für alle Bedürfnisse und Anwendungsfälle massenhaft Designer und Entwickler, die sich auf WordPress spezialisiert haben.

Auch die großen Webhoster kamen nicht umhin, auf den Zug aufzuspringen und Ihren Kunden WordPress direkt und möglichst einsteigerfreundlich anzubieten – obwohl sie damit ihren eigenen Website-Baukästen Konkurrenz machen. Falls Sie in Betracht ziehen, für Ihre Website einen solchen Baukasten zu nutzen, sollten Sie auf jeden Fall zunächst ein gehostetes WordPress-Angebot ausprobieren. Diese sind ab ein paar Euro im Monat zu haben und bieten oft mehr als die geschlossenen Systeme einzelner Hoster. Und wenn Ihnen Ihr aktueller Hoster nicht zusagt, können Sie mit WordPress – sechstens – Ihre Website-Inhalte mitnehmen und recht leicht zu einem anderen WordPress-Hoster umziehen.

Ganz ohne Einarbeitung geht es zwar auch bei WordPress nicht,

aber dank der aufgeräumten Bedienoberfläche kommt man schnell zu brauchbaren Ergebnissen. Dabei soll Ihnen der Artikel auf der nächsten Seite helfen. Im Beitrag ab Seite 68 stellen wir Ihnen einige vielfältig einsetzbare und schicke Themes vor und beschreiben, woran Sie gute Themes erkennen. Und der Artikel ab Seite 72 präsentiert Ihnen eine Auswahl besonders nützlicher Plug-ins. (*jo@ct.de*)



Mit der aufgeräumten Bedienoberfläche von WordPress kommen auch Menschen zurecht, die nicht fließend CSS und HTML sprechen.

Das Powertool R – hands-on!

Das Powertool R – hands-on! – websiteboosting.com

Da der Beitrag von Tobias Aubele in der letzten Ausgabe (Datenabzug von Google Analytics über das Tool R) auf große Resonanz gestoßen ist, möchten wir all denjenigen hier, die

noch keinen Kontakt zu dem kostenlosen Tool R hatten bzw. ihn bisher gemieden haben, eine kurze und leicht verständliche…

Da der Beitrag von Tobias Aubele in der letzten Ausgabe (Datenabzug von Google Analytics über das Tool R) auf große Resonanz gestoßen ist, möchten wir all denjenigen hier, die noch keinen Kontakt zu dem kostenlosen Tool R hatten bzw. ihn bisher gemieden haben, eine kurze und leicht verständliche Einführung geben.

Keine Sorge, es wird in diesem Beitrag nicht zu technisch. Und die Beispiele sind so gehalten, dass Sie einen schnellen und effizienten Einblick in das Tool bekommen, sofern Sie nicht nur lesen, sondern sich einfach selbst mal an die Tasten setzen. Wie bei vielen mächtigen Tools ist es oft die Anfangshürde der Bedienung, an der man scheitert. Gelingt es, diese zu überwinden, eröffnet sich wie so oft ein wahres Eldorado an Vereinfachungen für die eigene Arbeit und/oder zusätzliche Möglichkeiten für nötige Analysen oder die Gewinnung wichtiger Erkenntnisse.

Eigentlich ist R als Statistiktool konzipiert und bekannt. Aber im Lauf der Jahre ist durch eine schier unendliche Anzahl an Funktionsbibliotheken die Anwendungsbreite förmlich explodiert. So kann man z. B. Ergebnisse automatisch in HTML-Webseiten übertragen oder Textmining betreiben. Zwei zentrale Stärken von R sind wohl zum einen die Anzahl der Daten, die man damit sehr schnell verarbeiten kann. Zum anderen kann man aber alles, was man für ein Projekt oder eine Analyse experimentell durchgeführt hat, auf Knopfdruck ganz oder teilweise mit anderen Daten wiederholen. Richtig eingesetzt, kann das gerade im Online-Marketing sehr viel Zeit und damit Kosten sparen. Natürlich sind Sie nach der Lektüre des Beitrags kein R-Spezialist – aber dann können Sie für sich entscheiden, ob Sie dem Tool eine Chance geben und sich etwas tiefer darin eingraben. Im Web gibt es viele Tutorials, bei Udemy einen deutschen Kurs dazu und am Ende dieses Beitrags finden Sie Literatur für den Einstieg und für Fortgeschrittene zum Thema Data Science.

Erwarten Sie persönlich, dass die Menge der Daten künftig an

Ihrem Arbeitsplatz eher zunehmen wird? Wird die Zahl der Datenquellen eher größer und haben diese wahrscheinlich unterschiedliche Strukturen? Schreitet die Digitalisierung bei Ihnen gut voran? Dann, ja dann kann es sich wirklich lohnen, sich frühzeitig mit so einem Tool auseinanderzusetzen, das Sie bei der Bewältigung gut unterstützt. Und wie immer gilt: Wer früher dran ist, kann den anderen die lange Nase zeigen, während die sich wundern, wie man das Kaninchen mal wieder aus dem (neuen) Hut gezaubert hat …

Was ist R überhaupt?

R selbst ist eigentlich eher eine Programmiersprache als ein Tool und ähnelt Python. Über das ebenfalls kostenlose RStudio erhält R eine komfortable und übersichtliche Benutzeroberfläche, die Einsteigern etwas den Schrecken nimmt. Ursprünglich für statistische Anwendungen gedacht, ist R mittlerweile sehr viel mehr und hat große Stärken, wenn man Daten auswerten, umwandeln oder visualisieren muss. Sie möchten wissen, wie oft das Wort „Westernstiefel“ auf einer bestimmten Webseite vorkommt? Kein Problem – mit den entsprechenden kostenlosen Erweiterungspaketen (Librarys) ist das mit einer Befehlszeile erledigt. Sie brauchen einen Überblick über die Struktur der Shop-Umsatzdaten des letzten Jahres? Dazu laden Sie die Umsatzdatei ein und der Befehl „sum“ gibt Ihnen im Bruchteil einer Sekunde statistisch relevante Informationen. Vorausgesetzt, Sie wissen, was ein Median oder eine Standardabweichung ist. Mit anderen Worten sehen Sie mit einer Zeile Code, ob z. B. der häufig falsch verwendete Mittelwert als Kennzahl bei Ihren Daten überhaupt aussagekräftig ist.

> *„Windows, Mac oder Linux? Geht alles!“*

Wer tiefer einsteigt, kann mit R auch Anwendungen für Machine Learning fahren oder Data Mining betreiben, indem man im einfachsten Fall über die Daten Regressionsanalysen laufen

lässt. Bereits mit zwei bis drei Befehlszeilen kann man sich durch Umwandlung von Daten einen visuellen Überblick über mögliche Abhängigkeiten verschaffen. Ist der Umsatz tatsächlich vom Wetter abhängig und wenn ja, wie stark? Wie wirken sich Preisveränderungen von Mitbewerbern auf den eigenen Absatz aus und wie eine Erhöhung des Ads-Budgets? Gibt es Abhängigkeiten, die man bisher noch gar nicht auf dem Schirm hatte und wo es sich lohnt, tiefer zu graben?

R und RStudio sind für Windows, Mac OS und Linux verfügbar.

Was ist das Besondere an R?

Wie schon erwähnt, machen die vielen Funktionsbibliotheken (Librarys) R erst richtig nützlich und breit anwendbar. Diese lassen sich sehr simpel bei Bedarf einbinden und stehen fortan im Tool zur Verfügung. Damit kann man dann z. B. bestimmte Dateiformate lesen und erzeugen, Daten direkt in HTML-Vorlagen ausgeben, Anbindungen an verschiedene Datenbanken managen, linguistische Analysen durchführen oder auch Marktforschung betreiben. Und das Ganze völlig kostenlos, sehr stabil und performant.

Eine Besonderheit gegenüber anderen Programmiersprachen liegt darin, dass man in R mit sog. Vektoren arbeitet. Das erleichtert den Umgang mit Daten und es sind nicht wie sonst gesondert programmierte Schleifen nötig. Einen Vektor kann man sich als komplette Zeile (oder Spalte) in Excel vorstellen, der sich direkt verarbeiten lässt. Im Prinzip lässt sich so z. B. der komplette Quellcode einer Webseite in einen Vektor einlesen und über den Namen des Vektors ansprechen.

> *„R ist kostenlos, stabil und performant.“*

Wie Sie später noch sehen werden, ist es recht einfach, alle Befehle, die man in die Console eingegeben hat, später wieder aufzurufen und ggf. bei Bedarf in Summe in einen individuellen

Programmcode zu überführen. Das erleichtert das Verständnis, die Erstellung eines kleinen Programms und auch die Fehlersuche. Hier greift das gleiche Grundprinzip wie beim Arbeiten mit Excel, wenn man kein ausgebildeter bzw. fachkundiger Programmierer ist. Man bearbeitet Daten Schritt für Schritt und sieht jeweils sofort das Ergebnis und ob alles wie gewünscht passt. Nach Eingabe einer fehlerhaften Formel erscheint sofort eine Fehlermeldung – nicht erst, wenn später ein Programm läuft. Erst wenn alles passt, speichert man sich die einzelnen Befehlszeilen ab und lässt später direkt alle auf einmal ablaufen.

R und RStudio enthalten eine umfassende Hilfe: Durch Eingabe eines Fragezeichens, gefolgt von einem Funktionsbefehl wird die Hilfe aufgerufen, die die Verwendung genau erklärt. Auch im Reiter „Help“ (Abbildung 2; Ziffer 8) ist ein kleines Suchfeld integriert, das die Suche erleichtert. Empfehlenswert für einen strukturierten Einstieg ist es allerdings, in ein entsprechendes Fachbuch zu investieren. Eine kleine Auswahl finden Sie am Ende dieses Beitrags.

„R ist sehr flexibel bei der Datenausgabe.“

R hat sehr flexible Ausgabeformate. So können Ergebnisse, Tabellen, Grafiken und anderes über Erweiterungspakete z. B. als CSV, TXT, XLS, Word, PowerPoint oder auch direkt in HTML ausgegeben werden (Abbildung 1). Ebenso ist es natürlich möglich, die Daten in angebundene Datenbanken zu schreiben. Damit entfällt der sonst übliche Anwendungsbruch, weil man z. B. in Excel arbeitet, Diagramme erzeugt und diese dann für eine Präsentation in PowerPoint kopiert. Wer seine Berichte modern lieber gleich per HTML ins Intranet oder ins Web schießen möchte, kann das ebenfalls tun. Analyse, Aufbereitung und Dokumentation können also ein einem einzigen Tool erledigt und alle nötigen Schritte dazu reversibel zentral mit abgespeichert werden.

Abbildung 1: Auszug für gängige Ausgabe- bzw. Exportformate in R

„Nichts geht mehr verloren – kein einziger Arbeitsschritt.“

Vielleicht ist einer der größten Vorteile von R, dass Sie alles zusammen in einer Umgebung halten. Wie laufen üblicherweise Analysen im weitesten Sinne in Unternehmen ab? Man hat ein Set von Daten in verschiedenen Formaten und nicht selten strukturell unterschiedlich. Man kopiert bzw. liest diese dann oft in Excel ein, wandelt bestimmte Daten um, verdichtet, sortiert, filtert und modifiziert sie, führt weitere Konsolidierungen und Berechnungen durch und erzeugt diverse Diagramme oder Übersichtstabellen. Werden diese in einen Bericht gegossen oder präsentiert, kopiert man sie in Word oder PowerPoint. So weit, so gut. Im nächsten Monat/Quartal/Jahr oder beim nächsten Projekt steht dann eine Wiederholung des Ganzen an. Dabei entsteht oft das Problem, dass man nicht nur all diese Arbeiten erneut durchführen muss. Noch öfter steht man vor den fertig modifizieren Tabellen und weiß nicht mehr, wie man von den Ursprungsdaten dort hingelangt ist, da nicht mehr alle Formeln vorhanden sind bzw. neu erdacht und ausprobiert werden müssen. Alles in allem eine gigantische und auch ärgerliche Verschwendung von Zeit und Ressourcen.

Wenn Sie sich darauf einlassen, R etwas näherzukommen, gehört dies künftig der Vergangenheit an. Denn in R speichern Sie bei Bedarf jeden einzelnen Schritt – das Öffnen der Datei(en), jede Bearbeitung und schließlich auch die Erzeugung von Ergebnissen, Kennzahlen, Diagrammen und visuellen Auswertungen. Alles wird in einer Umgebung abgelegt und kann jederzeit nachvollzogen und reproduziert werden. Wenn Sie tiefer in R eintauchen, werden Sie sogar in der Lage sein, jedwede Auswertung automatisch per HTML oder PowerPoint neu zu erzeugen. Im einfachsten Fall kopieren Sie eine immer gleich benannte Datei mit den aktuellen Monatszahlen in das

Arbeitsverzeichnis von R und starten Ihre aufgezeichnete Programmierung. Das war es dann auch schon. Ein kompletter und durchaus auch optisch anspruchsvoller Bericht steht Sekunden später vollautomatisch im Intranet.

Wo finde ich R und RStudio?

Die Installation von R ist recht einfach. Es lässt sich kostenlos auf der Website von

cran.r-project.org

herunterladen, und zwar als Windows-, Mac-OS- und Linuxversion. R selbst ist wie erwähnt nicht so komfortabel zu bedienen und daher empfiehlt es sich für die meisten Nutzer, gleich noch RStudio dazu zu installieren. RStudio muss *nach* R installiert werden, sucht dann aber automatisch nach der R-Installation und integriert das Tool in eine übersichtlichere Benutzeroberfläche mit vielen nützlichen zusätzlichen Funktionen. Die Basisversion ist ebenfalls frei und unter rstudio.com erhältlich:

einfach.st/rstudio.

Wer den entsprechenden Rechner hat, sollte sich gleich die 64-Bit-Version installieren bzw. verwenden, da gerade rechenintensive Analysen deutlich schneller sind und auch die üblichen Restriktionen beim Datenvolumen unter 32 Bit entfallen.

Für normale Anwendungen ist RStudio kostenlos. Es gibt allerdings auch kommerzielle Lizenzen mit Support und weiteren Features sowie eine Serverversion.

Beim ersten Start prüft RStudio bzw. R noch, ob auf den Rechner eine aktuelle Java-Version installiert ist, was aus Sicherheitsgründen sowieso auf jedem Rechner Pflicht sein sollte. Anschließend präsentiert sich R integriert in RStudio aufnahmebereit für die ersten Befehle bzw. Gehversuche.

Abbildung 2: Der Startbildschirm von RStudio

Legen Sie jetzt einfach los

Nach der erfolgreichen Installation kann man sofort mit R bzw. RStudio arbeiten (Abbildung 2). Im linken Teil befindet sich die Console (Ziffer 1), in der die Befehle eingetippt werden. Geben Sie nach dem Promt (der blinkende Cursor nach dem „>“ Zeichen) einfach

2+5

ein und drücken Sie Return (Ziffer 2). Als Antwort erhalten Sie das Ergebnis dieser einfachen Rechenoperation. Gibt man einen Befehl falsch ein, z. B. „Print“ mit einem Großbuchstaben, erscheint eine Fehlermeldung, die hier zur Demonstration erzwungen wurde. Normalerweise erscheint nämlich bereits beim Tippen ein Vorschlagfenster mit den gültigen Befehlen für Funktionen (Ziffer 4). Das bedeutet, man kann sich den Rest einer Anweisung wie z. B. „print.data.frame“ sparen und den gewünschten Befehl einfach übernehmen.

Tipp

Wenn der Cursor im Consolenfenster blinkt, können Sie mit den Cursortasten hoch und runter die letzten eingetippten Befehle holen bzw. in diesen blättern. Das hilft insbesondere bei Tippfehlern, damit nicht alles erneut eingegeben werden muss. Einfach mit der Cursortaste-oben den letzten (falschen) Befehl holen, an der entsprechenden Stelle ausbessern und mit Return erneut auslösen. Zudem zeichnet R jeden Befehl in einer Historie auf, sodass er jederzeit auch später einfach noch mal verwendet werden kann.

Variablen weist man mit „<-“ ganz einfach einen Wert zu. Das machen wir hier nur einmal beispielhaft, um die Vorzüge des

RStudios zu zeigen. Ziffer 5 zeigt eine solche Zuweisung:

```
wsb <- 100
```

weist der Variable „wsb“ damit den Wert 100 zu. Also erst der gewünschte Name, dann die Zuweisungszeichen und anschließend der Wert. Im rechten oberen Teil von RStudio findet man dann unter dem Reiter „Environment“ (Ziffer 6 und das angezeigte Fenster mit Ziffer 6) genau diese und natürlich alle anderen Variablen wieder, die man im Lauf einer Sitzung definiert hat. Das dient der Kontrolle, welchen Wert eine Variable gerade hat. Wenn Sie jetzt als Befehlszeile

```
wsb + 5
```

gefolgt von Return eingeben, sehen Sie unter Environment rechts oben den neuen Wert 105 für diese Variable. Prinzipiell kann man ganzen Datenpools (Vektoren, Matrizen, Data Frames) Variablennamen zuweisen, was das Handling und die Lesbarkeit durch einfache Namensvergabe enorm vereinfacht. Dies aber nur als Hinweis.

Unter dem Reiter „History“ (Ziffer 3) werden alle Befehle aufgezeichnet, die man in die Console eingegeben hat. Mit einem Mausklick lassen sie sich einfach wiederholen bzw. in die Console übernehmen. Später lassen sich diese Befehle dann ganz einfach in einem kleinen Programmcode überführen, der nacheinander abläuft.

Das charmante Prinzip ist also für den Einsteiger, dass er zunächst Schritt für Schritt nach jeder Eingabe explorativ prüfen kann, ob alles so richtig ist, und erst danach fasst man quasi die einzelnen Befehle zu einem ablauffähigen Programm zusammen und kann dieses dann mit einem einzigen Befehl starten.

Einfach mal machen!

Alle hier gezeigten Beispiele sind so gehalten, dass Sie möglichst einfach selbst erste Schritte zum Ausprobieren unternehmen können. Wir haben daher oft auf branchenübliche Zahlenbeispiele verzichtet, weil R einige Datenquellen zum Experimentieren bereits mitliefert. Sie müssen sich also nicht erst mit dem Erstellen und Einlesen von Datensätzen herumschlagen, nur um überhaupt eine Grundlage zum Testen zu haben. Ob Sie nun die Kelchlänge von Lilienarten oder den Benzinverbrauch von Autos visualisieren oder Marketingausgaben oder Backlinks, ist von der Methode her ja austauschbar. Hier liegt der Fokus tatsächlich darin, Sie sofort in die Lage zu versetzen, einfach einmal mit dem Tool herumzuspielen.

Ein Anwendungsbeispiel: Sie nutzen regelmäßig einen oder mehrere Datensätze, z. B. aus einem SEO-Tool, mit denen Sie arbeiten müssen. Diese enthalten aber unerwünschte (Sonder-)Zeichen, wie Umlaute als &szlig“ statt einem „ß“ oder schlicht den typischen Punkt statt des deutschen Kommas als Kommastelle. Das kann man natürlich auch in Excel erledigen, aber man muss alle Schritte einzeln und jedes Mal durchführen. Über R lädt man sich so einen Datensatz, verändert ihn mit den entsprechenden Anweisungen einmalig und prüft, ob alles korrekt ist. Anschließend speichert man sich die Befehle einfach ab und kann sie später direkt aufrufen. Der Datensatz muss dann nur noch geladen werden, das entsprechend passende erstellte Programm wird gestartet und der Datensatz entweder manuell oder sogar schon vom eigenen Programm abgespeichert – fertig. Prinzipiell lassen sich damit alle Daten auf einfache Weise verändern, bereinigen, auf Konsistenz prüfen oder einfach nur einzelne Zeichen tauschen.

Im rechten unteren Bereich von RStudio (Ziffer 8) hat man Zugriff auf Datenfiles, Grafiken (Plots), Packages (Erweiterungen) sowie auf eine Hilfe und einen Viewer.

Abbildung 3: In RStudio findet man unter dem Menüpunkt „Help“ nützliche Cheatsheets

Abbildung 4: In R selbst (nicht in RStudio) sind umfangreiche PDF-Handbücher hinterlegt (englisch)

Mitgelieferte Testdateien

Praktischerweise beinhaltet die Installation von R bereits einige Dateien mit Daten, sodass man die ersten Gehversuche unternehmen kann, ohne erst einmal selbst Daten einlesen zu müssen. Eine dieser Beispieldatensätze ist „state.x77“. Er lässt sich mit dem Befehl

```
View(state.x77)
```

anzeigen. Links oben im Fenster werden jetzt alle 50 US-Staaten mit einigen Spalten wie „Einwohnerzahl“, „Durchschnittseinkommen“ etc. angezeigt (Abbildung 5).

Abbildung 5: R liefert einige Beispieldaten automatisch mit

Ein einfacher Befehl zeigt Ihnen direkt statistisch relevante Daten dieses Datensatzes an (Abbildung 6):

```
summary(state.x77)
```

R zeigt für jede Spalte gesondert die statistisch wichtigsten Strukturinformationen an. Für eine spätere Datenanalyse oder die Bildung von Kennzahlen ist dies sehr nützlich, weil man sofort u. a. die Streuung ablesen kann. Ein extremes Beispiel zur Verdeutlichung: Hatten 200 Verkäufe den Wert 1 € und weitere 200 den Wert 100 €, dann macht die Verwendung eines Mittelwerts (50 €) nicht wirklich Sinn.

Abbildung 6: Statistisch relevante Strukturdaten der Datei „state.x77″

Der Befehl

```
state.x77[11,2]
```

greift z. B. direkt im Datensatz „state.x77“ auf die elfte Zeile und die zweite Spalte zu und liefert 4963 als das Durchschnittseinkommen in Hawaii zurück (in Abbildung 5 zu sehen). Genauso können über Spalten oder Zeilen natürlich entsprechende weitere einfache Berechnungen oder Analysen durchgeführt werden.

Die ebenfalls in R mitgelieferte Datei „state.name“ beinhaltet nur die Namen der 50 US-Staaten. Welche dieser Namen besteht aus mehreren Wörtern wie z. B. New Jersey? Hier hilft die Funktion „grep“, mit der man in diesen Daten ganz einfach nach einem Leerzeichen sucht. Beachten Sie bitte, dass sich der Name der verwendeten Datendatei jetzt ändert bzw. eine andere Datei (state.name) zugrunde liegt (Testen Sie doch mal View(state.name) um zu sehen, was der Datensatz beinhaltet.)

```
state.name[grep(“ „, state.name)]
```

Als Ergebnis erhält man in der Ausgabe die zehn Staaten mit einem „Doppelnamen“. Tauschen Sie das Leerzeichen z. B. gegen den Wortbestandteil „New“ aus, erhalten Sie die vier Bundesstaaten, die eben diese Zeichenfolge enthalten, der Befehl lautet dann:

```
state.name[grep(„New“, state.name)]
```

Diese einfachen Beispiele sollen nur verdeutlichen, wie leicht die Extraktion und Filterung in Datenbeständen ist und welches Prinzip dahintersteckt. Natürlich kann man solche Aufgaben bei so kleinen Datensätzen auch gut und einfach z. B. in Excel erledigen. Werden die Daten jedoch umfangreicher, macht das schon deutlich mehr Arbeit, die zudem prinzipiell jedes Mal neu durchgeführt werden muss. Die Befehle bzw. Befehlssequenzen hier in R können jedoch abgespeichert und zu jedem späteren Zeitpunkt einzeln oder in Summe aufeinanderfolgend gestartet werden.

Tipp: Geben Sie den Befehl

datasets::

ein und Sie erhalten in einem Pop-up-Fenster die in R integrierten Versuchsdatensätze angezeigt, die Sie per Mausklick einfach übernehmen können (Abbildung 7). Geben Sie den Befehl

data()

ein, erscheinen im linken oberen Fenster alle Namen und eine jeweils kurze Beschreibung aller Datensets, die bereits installiert sind.

Abbildung 7: Automatisch mitgelieferte Datensätze zum Ausprobieren mit Inhaltsbeschreibung

Noch ein weiteres kleines Beispiel zeigt Abbildung 8. Für die Variable „Zeichenkette“ wird zunächst hilfsweise ein Text „Die Website Boosting …“ eingelesen. Der Befehl „strsplit“ zerlegt dann den Text in einzelne Wörter bzw. trennt nach einem Leerzeichen. Stellen Sie sich vor, Sie hätten eine Exceltabelle mit mehreren Tausend Zellen, die jeweils den Text einer Webseite enthalten, und für semantische Analysen müsste jedes Wort extrahiert werden. In R liest man dazu vereinfacht erklärt über ein Erweiterungspaket ein Set von URLs mit einer einzigen Funktion in Vektoren ein und ein weiterer Befehl extrahiert aus jedem Vektor jedes einzelne Wort. Bereits mit einigen Grundkenntnissen in R lässt sich so durchaus enorm Zeit sparen.

Unter einfach.st/rdatasets finden Sie übrigens eine recht gut kommentierte Übersicht (englisch) mit Beispielen und Beispielcode für die mitgelieferten Data Sets.

Abbildung 8: Mit einem einzigen Befehl Texte in Wörter zerlegen

Erweiterungspakete installieren

Wie erwähnt werden für R sehr viele Erweiterungspakete im Web angeboten, die sich recht einfach integrieren lassen. Diese muss man nur einmal zu R hinzuladen und kann sie später dann nach dem jeweiligen Programmstart bei Bedarf aktiveren und die dort hinterlegten Funktionen nutzen. Eine erste gute Quelle finden Sie auf cran.r-project.org. Die dort ladbaren Pakete kann man direkt über R laden. Dafür muss bzw. sollte man natürlich den Namen der Erweiterung kennen. Diese werden meist in Anleitungen, Beispielen oder in der Literatur erwähnt. Die Installation kann dann per Befehlszeile über die Funktion „install.packages“ erfolgen (hier die Erweiterung „htmltidy“ zum Umgang mit HTML-Dokumenten):

```
install.packages(„htmltidy“)
```

oder über das Menü oben in RStudio unter Tools/Install Packages, wie in Abbildung 9 zu sehen ist. Auch hier erscheint beim Tippen bereits ein kleines Pop-up mit einer entsprechenden Auswahl. Lässt man den Haken „Install depencies“ (in der Abb. verdeckt unter dem Pop-up) drin, werden sog. abhängige Packages gleich mit installiert. Die Erweiterung „htmltidy“ greift z. B. unter anderem auf die Erweiterung „xml“ und „htmlwidgets“ zu und löst damit auch gleich deren automatische Integration aus. In der Console erscheint dann eine entsprechende Meldung.

Tipp

Manchmal zickt R bei der Eingabe von Erweiterungspaketen über die Kommandozeile mit einer Fehlermeldung zurück, weil die Groß-/Kleinschreibung nicht passt. Daher empfiehlt es sich tatsächlich, das (einmalige) Laden von Packages – aus Sicht eines Programmierers sicher unehrenhaft, aber eben einfacher – über das Menü von RStudio wie in Abbildung 9 gezeigt zu erledigen.

Die meisten Packages sind recht gut dokumentiert. So findet man für das eben erwähnte Package weitere Infos unter cran.r-project.org/web/packages/htmltidy/index.html und dort unter „Downloads“ auch ein PDF mit einer genauen Funktionsbeschreibung.

Möchte man die Funktionen eines einmal installierten Packages in R verwenden, aktiviert man es ganz einfach mit der Anweisung:

library(packagename) – im Beispiel also: library(htmltidy)

Abbildung 9: Erweiterungen lassen sich sehr einfach installieren

Einen Überblick über installierte Erweiterungen findet man in RStudio im rechten unteren Fenster unter dem Reiter „Packages“ (siehe Abbildung 2, Ziffer 8).

Eine CSV-Datei einlesen, modifizieren und wieder ausgeben

Zum Einlesen von Dateien ist es nützlich, gleich zwei Erweiterungen zu installieren: „readr“ und „readxl“ (für das Excelformat). Dazu geben Sie einfach die beiden Befehlszeilen ein

install.packages(„readr“)

install.packages(„readxl“)

und aktivieren anschließend zumindest die Erweiterung „readr“:

library(readr)

Erstellen Sie zum Ausprobieren eine einfache CSV-Datei z. B. über Excel oder einen Texteditor mit mehreren Zeilen und Spalten. In die erste Zeile schreiben Sie durch Kommata getrennt die Spaltenüberschriften und darunter ebenfalls durch

Kommata getrennt die Datensätze, wie z. B.

```
URL,Statuscode,Inlinks
www.meine-domain.de,200,1.250
www.deine-domain.com,200,3.243
www.ihre-domain.at,200,13.283
```

Der Punkt bei den Inlinks-Zahlenwerten ist absichtlich gewählt, er soll durch ein Komma getauscht werden, um das Prinzip der Datenmodifikation zu zeigen. Erzeugen Sie in RStudio über „File“ – „New Projekt“ ein neues Projekt und speichern Sie dieses auf dem Rechner unter einem beliebigen Ort ab. Der angegebene Projektname ist dann automatisch der Verzeichnisname und Ihr „Working Directory“. In dieses Verzeichnis speichern Sie dann die CSV-Datei ab, damit Sie diese durch einen einfachen Befehl öffnen und später speichern können. Alternativ kann man mit R natürlich auch auf alle anderen Verzeichnisse/Dateiorte zugreifen, aber die Dateien zu Anfang in einem eigenen Working Directory zu speichern, erleichtert die Arbeit etwas. In dem Beispiel hier wurde als Working Directory „WSB Einführungsartikel“ gesetzt und rechts unten unter „Files“ taucht die Datei beispiel.csv nach der Erzeugung bzw. dem Abspeichern dann auch entsprechend auf (Abbildung 10).

Der Befehl

```
getwd()
```

gibt übrigens das aktuelle Verzeichnis aus.

Abbildung 10: Legen Sie über ein neues Projekt einen Speicherort zum Testen an

Über den Befehl

```
beispiel <- read.csv(„beispiel.csv“)
```

weisen Sie jetzt ganz einfach der Variable „beispiel“ (oder auch einem beliebigen anderen Namen) über die Funktion

„read.csv“ die Datei beispiel.csv zu. Da die Datei im Working Directory liegt, brauchen Sie keine gesonderten Pfadangaben. Den Inhalt dieser Datei kann man anschließend einfach über den Variablennamen ansprechen. Tippen Sie z. B.

```
print(beispiel)
```

ein, wird der Inhalt der eingelesenen Datei ausgegeben (Abbildung 11). Probieren Sie auch den Befehl view(beispiel) und statt der Ausgabe in der Console geht links oben ein Datenfenster auf, das die Daten sauber in Tabellenform anzeigt.

Abbildung 11: Dateien einlesen und anzeigen lassen

Geht es nicht nur um Datenanalysen, sondern werden Daten auch modifiziert oder neue Daten erzeugt, die man speichern möchte, geht auch das recht einfach mit dem Befehl write.table:

```
write.table(datensatz, „dateispeichername.csv“)
```

Möchten Sie z. B. die Daten des integrierten Beispiels „state.x77“ in eine CSV-Datei namens „test.csv“ in das aktive Arbeits-/Projektverzeichnis speichern, lautet der Befehl also:

```
write.table(state.x77, „test.csv“)
```

Abbildungen erzeugen

Natürlich lassen sich Grafiken in Excel recht einfach erzeugen. R hält jedoch deutlich mehr Möglichkeiten und nach Meinung vieler auch bessere Grafiken vor. Wer z. B. schon einmal versucht hat, aus einem Datensatz ein vernünftiges Histogramm zu erzeugen, wird sich über die fehlende Möglichkeit geärgert haben, die automatische Clusterung von Excel zu umgehen. Hat man stark abweichende Werte, sind Histogramme dort praktisch wegen der Spreizung nicht zu verwenden. In R stellt das Definieren von Clustern für die Histogrammsäulen kein Problem dar. Ebenso wenig wie die

(befehlsorientierte) Formatierung einer Abbildung. Einmal definiert, lässt sie sich für Folgeabbildungen mit einer Zeile Code in immer demselben eigenen Design erzeugen.

Abbildung 12: Anriss des mitgelieferten Datensatzes „mtcars“

Ein Histogramm erzeugt man in R mit dem einfachen Befehl „hist“ gefolgt von dem gewünschten Datensatznamen in Klammer dahinter. Eine der mitgelieferten Datenbeispiele ist die Datei mtcars. Mit dem nun schon bekannten Befehl

```
print(mtcars)
```

können Sie einen schnellen Blick darauf werfen. Möchte man nun den Benzinverbrauch aller dort gelisteten Modelle (erste Spalte, mpg, also miles per gallon) in einem Histogramm darstellen, benötigt man nur eine Befehlszeile:

```
hist(mtcars$mpg, col = „blue“)
```

Das Prinzip ist recht leicht ersichtlich. Aus den Daten „mtcars“ soll aus der Datenspalte „mpg“ ein Histogramm erzeugt werden (Befehl „hist()“), dessen Balken blau sind („blue“). Das Ergebnis zeigt Abbildung 13 auf der rechten Seite – und im Vergleich links das Histogramm, das Excel mit den gleichen Daten erzeugt. R teilt die Datencluster statistisch vernünftiger auf und erstellt ein durchaus differenziertes Bild. Wenn Sie die Clusterbildung selbst beeinflussen möchten, weil Sie z. B. feinere Abstufungen brauchen oder einfach nur auch zeigen wollen, dass unter zehn Gallonen pro Meile eben kein Wert und damit keine Säule vorhanden ist, geht das über eine einfache Erweiterung des Befehls mit Angabe der gewünschten Intervalle „breaks“ (der Übersichtlichkeit halber wurde auf die Farbangabe hier verzichtet). Probieren Sie es einfach aus:

```
hist(mtcars$mpg, breaks = c(5,10,15,20,25,30,35))
```

Selbstverständlich könnte man die Intervalle auch mit einer

eigenen Formel statt mit manuell eingegebenen Intervallen wie hier im Beispiel berechnen lassen.

Abbildung 13: Vergleich eines unbearbeiteten Histogramms mit identischen Daten in Excel und R

Aber auch ein schneller optischer Blick auf strukturelle Eigenheiten von Datenreihen ist möglich. Dazu lässt man sich eine oder mehrere Datenreihen als Boxplot ausgeben. Zur Demonstration kann man ganz einfach erneut auf das bereits genutzte Datenset „mtcars“ zugreifen. Möchte man mehr über den Einfluss der Anzahl der Zylinder eines Motors auf den Verbrauch bzw. auf die Streuung der Daten wissen, setzt man beides einfach in Beziehung. Der Befehl dazu ist

```
boxplot(mpg ~ cyl, data=mtcars)
```

In Abbildung 12 ist ersichtlich, dass es zwei Spalten im Datenset gibt, die mit „mpg“ (Miles per Gallon) und „cyl“ (Zylinder) überschrieben sind. Der Befehl „boxplot“ bezieht sich auf die angegeben Datenquelle, setzt die beiden Datenreihen dann ganz einfach in Beziehung und gibt Abbildung 14 aus.

Abbildung 14: Statistische Kennwerte visualisieren

Dort erkennt man auf den ersten Blick, wie stark die Verbrauchsdaten des nach Zylinder geclusterten Verbrauchs streuen. Würde man statt mtcars auf einen eigenen eingelesenen Datensatz verweisen und zwei Spaltenüberschriften auswählen, müsste man den boxplot-Befehl einfach nur entsprechend anpassen und mehr nicht. Dazu wären in R nur zwei Befehlszeilen nötig – der zum Einlesen und der Befehl zur entsprechenden Ausgabe. Eine Sache von wenigen Sekunden.

Kleines Cheatsheet für statistische

Funktionen

Durchschnitt/Mittelwert: mean()
Median (zentraler Wert): median()
Varianz: var()
Standardabweichung: sd()
Quantile: quantile()
Bandbreite von–bis: range()

> *„Vor der Verwendung von Daten ist es immer hilfreich, einen statistisch motivierten Blick auf die Struktur dieser Daten zu werfen.“*

Wer es eine Spur härter, aber keinesfalls komplexer haben möchte, kann auch problemlos mehrere Datenreihen miteinander korrelieren lassen bzw. optisch über Streudiagramme prüfen, ob es einzelne Abhängigkeiten bzw. Korrelationen gibt. Dies lässt sich gut bzw. schnell und einfach am Beispiel des Datensets „iris“ zeigen. Dort sind für drei Schwertlilienarten in Summe 150 Datensätze mit jeweils Länge und Breite der Kelchblätter (Sepal) und der Kronenblätter (Petal) hinterlegt. Mit

View(iris)

wird das Datenset angezeigt (Abbildung 15).

Geben Sie jetzt doch einfach den Befehl

plot(iris[-5])

ein. Sie erhalten zwölf Streudiagramme aller möglichen Kombinationen zwischen Kelchblättern und Kronenblättern nach Länge und Breite (Abbildung 16). In den beiden rot markierten Diagrammen lässt sich unmittelbar ein fast linearer Zusammenhang bezüglich Länge und Breite bei den Kelchblättern erkennen, während bei den Kronenblättern und zwischen Kronen- und Kelchblättern kein solcher offensichtlich ist.

Abbildung 16: Korrelieren einzelne Daten miteinander?

Fazit

Tauschen Sie jetzt einfach geistig oder auch real in der Praxis die letzten Beispiele wie z. B. das florale Datenset mit echten Daten aus Ihrem Arbeitsumfeld aus – dann erkennen Sie zumindest in Ansätzen den Wert und die Transparenz, die R in wenigen Minuten bringen kann. Und das ganz ohne wirklich tiefergehende Programmierkenntnisse und nur mit dem Wissen über einige einzelne Befehlszeilen. Natürlich bleibt der Nutzen tatsächlich überschaubar, wenn Sie bei der Lektüre nur dieses Beitrags stehen bleiben. Aber stellen Sie sich doch mal vor, Sie steigen jetzt motiviert und in Erahnung der Kenntnis des Potenzials von R Stück für Stück tiefer ein. Sie haben noch nicht einmal ein halbes Prozent von dem ausprobiert, was R leisten kann, wenn Sie es richtig bespielen. Was nun am Ende mehr nützt – die Rationalisierungseffekte für einfacheres und schnelleres Arbeiten oder neue Erkenntnisse und bessere, wirklich datengestützte Entscheidungen –, bleibt Ihren Überlegungen und Ihrer Motivation überlassen.

Vitamin D: Datenanbindung über R an Google Schnittstellen

Vitamin D:Datenanbindung über R an Google

Schnittstellen – websiteboosting.com

R ist nicht nur Open Source (damit kostenlos) und anpassungsfähig durch vielfältige Pakete, sondern auch offen für andere (technische) Verbindungen. Diese Verbindungen, vulgo API, ermöglichen, flexibel Daten an der Ursprungsquelle abzufragen, entsprechend weiterzuverarbeiten und ggf. modifiziert…
R ist nicht nur Open Source (damit kostenlos) und anpassungsfähig durch vielfältige Pakete, sondern auch offen für andere (technische) Verbindungen. Diese Verbindungen, vulgo API, ermöglichen, flexibel Daten an der Ursprungsquelle abzufragen, entsprechend weiterzuverarbeiten und ggf. modifiziert wieder zurückzuliefern. Google bietet eine Vielzahl von API, die in R eingebunden eine enorme Individualisierung und Automatisierung von analytischen Prozessen und Visualisierungen ermöglichen. Die Excel-Alternative Google Sheets in Kombination mit Google Data Studio ist bereits eine unkomplizierte Lösung für vielfältige Visualisierungen. Gepaart mit der statistischen Leistung von R eröffnet sich durch die weitere Kombination mit API nochmals ein deutlicher Zugewinn an Flexibilität! Tobias Aubele zeigt, wie Sie diese Schnittstellen nutzen können.

R wird kontinuierlich von einer großen Community weltweit weiterentwickelt, Gleiches gilt für die Dienste von Google wie Maps & Co. Unglaubliche Mengen an Daten liegen in diversen Datensilos und „sehnen“ sich danach, verarbeitet bzw. interpretiert zu werden und einen Nutzen zu stiften. Durch die Vielzahl an Silos und steigenden Datenmengen bedarf es einer effektiven und effizienten Verarbeitung, damit letztlich genügend Zeit für die Interpretation der Daten und das Ableiten von Maßnahmen übrig bleibt. Obwohl ggf. sehr zeitaufwendig, wird in Fachbereichen gerne Microsoft Excel eingesetzt, da dort Daten mittels Formeln und Funktion aus diversen Quellen flexibel zusammengeführt und verarbeitet werden können. Leider sind Daten teilweise veraltet, ehe sie letztlich über manuelle Zusammenfassungen und Verdichtungen

für den eigentlichen Nutzungszweck zur Verfügung stehen. Sofern direkt API (Programmierschnittstellen) genutzt werden, können aktuelle Daten unmittelbar am Ursprung abgeholt und zurückgeliefert werden und bieten langfristig ggf. ein hohes Potenzial für automatisierte Abläufe.

Programmcode online verfügbar!

Damit Sie Codezeilen nicht einzeln abtippen müssen, sind zwei der Skripte aus diesem Beitrag direkt online für Sie zum Download hinterlegt:
Das Skript zum Knowlede Graph unter www.websiteboosting.com/code/r-knowledge
Das Skript für Google Sheets unter www.websiteboosting.com/code/r-sheets

Die API – aktuelle Daten direkt vom Ursprung

Über die Google API können diverse Google-Dienste direkt über R angesprochen werden. Nach einer Registrierung unter console.cloud.google.com kann die gewünschten API, wie bspw. die des Google Knowledge Graph, einzeln aktiviert und unmittelbar getestet bzw. genutzt werden (siehe Abb. 1).

Abb. 1: Aktivierung einer Google API
Weiterhin können anschließend die notwendigen Anmeldedaten wie ein API-Schlüssel oder eine OAuth-Client-ID erstellt werden (siehe Abb. 2). Damit wird sichergestellt, dass der Datenzugriff nur von autorisierten Nutzern ermöglicht wird.

Abb. 2: Erstellung von gesicherten Anmeldemechanismen
Mittels der API werden die Daten direkt aus der Datenbank von Google abgerufen. Es stehen damit nicht nur aktuelle, sondern auch ggf. erweiterte Daten zur Verfügung. Das heißt,

Informationen über Albert Einstein, welche im Graphen am rechten Bildschirmrand angezeigt werden (siehe Abb. 3), stehen direkt zur weiteren Verarbeitung zur Verfügung. Die API ermöglicht die Extraktion weiterer Daten, sollten mit dem Suchbegriff mehrere Entitäten (Personen, Objekte) in Verbindung gebracht werden wie bspw. die Person Hans Albert Einstein.

Abb. 3: Google Knowledge Graph von Albert Einstein

Das zur Datenextraktion notwendige R-Paket lautet *GoogleKnowledgeGraphR*, entwickelt von Daniel Schmeh (zur Dokumentation siehe das PDF unter: einfach.st/r4kgraph). Mittels des individuellen API-Keys (Generierung siehe Abb. 2) werden mit dem Befehl

```
gkg_raw
```

die gewünschten Informationen in diversen Sprachen abgerufen. Das heißt, es stehen neben der Beschreibung unmittelbar die URL des Bildes, die URL des Eintrages sowie Informationen zu den Nutzungsrechten (dennoch prüfen!) zur Verfügung (siehe Abb. 4). Ohne die jeweilige Suchabfrage bei Google starten zu müssen, können die Ergebnisse direkt in R abgerufen und weiterverarbeitet werden. Dies kann zu einer weiteren Vereinfachung und damit zur Zeitersparnis führen.

Abb. 4: Aufruf der Knowledge Graph API über R

Ähnlich diesem Prozess erfolgt die generelle Nutzung von Google API (mit R). Es bedarf eines API-Schlüssels bzw. einer Autorisierung sowie eines korrespondierenden Programms (hier R-Paket). In der Praxis können nahezu alle Google API, wie bspw. die Google Translator API oder die Google Maps/Places API, abgefragt werden. Vereinzelt sind die Services kostenpflichtig mit einem flexiblen Abrechnungsmodell wie bspw. auf Basis der Anzahl an API-Abfragen. Damit könnten Textübersetzungen bzw. Abfragen zu Distanzen, Öffnungszeiten,

Adressen etc. von Restaurants und anderen Google-Places-Einträgen vollzogen werden. Die entsprechenden R Packages heißen *googleway* und *googleLanguageR*.

Kollaboration mit Google Sheets

Das Besondere an Google Sheets ist der weltweite Zugriff, die große Rechenleistung sowie die gleichzeitige Zusammenarbeit mit mehreren Personen. Weiterhin können Daten aus diversen Quellen zusammengeführt (vergleichbare Excel-Funktionen SVERWEIS bzw. INDEX/VERGLEICH) und letztlich ein Bericht sehr flexibel gestaltet, individualisiert und geteilt werden. Daher erscheint es sinnvoll, die Möglichkeiten von Google Sheets sowie ggf. der nachgelagerten Visualisierung mit Google Data Studio zu nutzen.

Daten in R können an Google Sheets bzw. von Google Sheets übermittelt werden. Hierzu sind die Pakete *googlesheets* (PDF zur Dokumentation siehe: einfach.st/r4sheets) bzw. *googleAuthR* notwendig bzw. hilfreich. Nach der Installation des Paketes über den Befehl

install.packages(“googlesheets“)(siehe Abb. 6)

erfolgt die Freigabe bzw. Authentifizierung des R-Pakets mit Google Drive über

gs_auth() *(*siehe Abb. 5).

Abb. 5. Autorisierung des R-Pakets
Nach erfolgreicher Authentifizierung kann mit dem Befehl

gd_user()

der aktuelle Nutzer sowie der Berechtigungsstatus abgefragt werden. Der Befehl

gs_new

ermöglicht, ein neues Sheet anzulegen sowie den Namen bzw.

Arbeitsblattnamen zu vergeben. Ab diesem Zeitpunkt können Daten eingegeben bzw. das Sheet genutzt werden. Der Befehl

gs_delete

löscht das Sheet entsprechend. Das jeweilige Sheet muss vor einer Aktion immer exakt adressiert werden, wozu diverse Möglichkeiten zur Verfügung stehen (siehe Dokumentation). Die exakte Art ist über die URL via

gs_url (siehe Abb. 6),

sofern Namen einzigartig sind, auch hilfsweise über den Titel (Befehl: *gs_title*). Soll das Sheet im Browser aufgerufen werden, muss der Befehl

gs_browse

mit dem Namen des Sheets (im Beispiel in der Variable gs gespeichert) und dem Blattnamen (relative Nummer bzw. absoluter Name) ausgeführt werden.

Abb. 6: Erstellung eines Google Sheets innerhalb von R

Google Sheets mit Daten füllen

Prinzipiell können auch Daten aus Google Analytics in R abgerufen und analysiert werden (siehe den ersten Teil dieses Beitrags in der letzten Ausgabe #53). Dem Google Sheet „Analytics“ kann über den Befehl

gs_ws_new

ein neues Arbeitsblatt mit dem Namen Oktober hinzugefügt werden (siehe Abb. 7) und anschließend mit den Google-Analytics-Daten befüllt werden (Abruf der Daten über den Befehl *google_analytics* aus dem R-Paket googleAnalyticsR). Die neuen Informationen sollen ab der Position (anchor) A1 zeilenweise (byrow = FALSE) eingefügt werden. Sollten dann im späteren Verlauf weitere Zeilen hinzugefügt werden (und dabei

die alten Daten nicht überschrieben werden), kann ein Arbeitsblatt ausgelesen (Befehl *gs_read*), die aktuelle Anzahl an benutzten Zeilen festgestellt (nrow) und anschließend über die R-Funktion *paste0* ein neuer Start ermittelt (im Beispiel Zelle A38) und an die Variable gs*_position* übermittelt werden. Ein weiterer Aufruf der Funktion *gs_edit_cells* fügt die Daten demnach nicht wieder an der Position A1 ein, sondern nutzt als initiale Startposition den Inhalt der Variable gs_position.

Abb. 7: Daten an Google Sheets aus R übertragen

Highlight: spezifische Google-Sheets-Funktionen

Mittels Austausches von Information an Google Sheets (lesen und schreiben) kann ein flexibler Datenaustausch mit R erfolgen. Ein weiterer Vorteil von Sheets ist die Mannigfaltigkeit spezieller Funktionen. Über die Funktion

```
=IMPORTXML(VERKETTEN("http://www.meineSeite.de";A2);"//h1")
```

kann eine XPath-Abfrage generiert werden, welche Elemente aus beliebigen Webseiten direkt ausliest (hier der H1-Text; siehe Abb. 8). Mittels einer weiteren Funktion

```
=GOOGLETRANSLATE(D2;"de";"en")
```

wird der Inhalt einer Zelle in die englische Sprache übersetzt. Die Funktion Verketten ermöglicht in diesem Zusammenhang, URL-Fragmente aus Google Analytics zu einer kompletten URL zusammenzuführen. Die Daten in Google Sheets können im weiteren Verlauf (manuell) beliebig verändert und anschließend wieder in R eingelesen/ weiterverarbeitet oder mit einer Visualisierung in Google Data Studio verbunden werden.

Abb. 8: Nutzung von Google-Sheets-spezifischen Funktionen

Google Data Studio hat diverse Verbindungen zu Datenquellen, was für die Analyse und Aufbereitung eine hohe Flexibilität bietet. Weiterhin können beliebige Google-Tabellen als Datenquelle initiiert und anschließend in die Berichte integriert werden – bspw. jeweils als eigene Seiten. Die Anbindung geschieht über zwei Schritte. Erstens muss über den Google Tabellen-Connector das jeweilige Tabellenblatt gewählt und somit als Datenquelle definiert werden (siehe Abb. 9). Zweitens müssten anschließend die damit geschaffenen Datenverbindungen mit einem spezifischen Bericht verbunden und die jeweiligen Datenfelder ausgewählt werden. Zur Visualisierung stehen diverse Möglichkeiten zur Verfügung (siehe Abb. 10).

Abb. 9: Auswahl von Google-Tabellen als Data-Studio-Connector

Innerhalb des Berichtes besteht die Möglichkeit, mit mehreren Seiten zu arbeiten und damit bspw. die Sheets-Integration als Detailbericht einer großen Analyse zu nutzen. Weiterhin können einzelne Informationen/Spalten aus Sheets ausgewählt und in diversen Visualisierungsmöglichkeiten auf dem Bericht verortet werden.

Abb. 10: Berichtsaufbau mit Sheets-Integration

Fazit

Zusammenfassend lässt sich sagen, dass R, Google Sheets, Google Data Studio etc. für sich hervorragende Möglichkeiten bieten. Die wahre Stärke zeigt sich, wie fast immer, in der Kombination verschiedener Faktoren/Quellen. Daten aus diversen Quellen (schnell) zusammenzufassen, aus unterschiedlichen Blickwinkeln zu betrachten und unkompliziert einem definierten Empfängerkreis zur Verfügung zu stellen, sind nur punktuelle Vorteile.

In diesem Sinne: Verbinden Sie Ihre Informationen, kombinieren

Sie Daten – ggf. trügt ja der erste Anschein.

R4SEO: Automatisierte Reports mit R generieren (Teil 1)

R-Leuchtungen! Teil 9

R-Leuchtungen! Teil 9: Title-Geddon – war da was bei Ihnen? – websiteboosting.com

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…

In der letzten Ausgabe der Website Boosting gab Tobias Aubele einen sehr guten Einblick in die Potenziale von R, um auf einfache und reproduzierbare Weise Daten über die Google-Analytics-API abzufragen. Diese neue Artikelserie von Patrick Lürwer baut darauf auf und zeigt, wie mittels R Reports erstellt werden können, die sich zu einem bestimmten Zeitpunkt automatisch aktualisieren und per Mail verschicken. Im ersten Teil geht es nun zunächst um das Abfragen verschiedener APIs für die Datenbeschaffung. In den folgenden Teilen werden die Daten aggregiert, visualisiert und in einem Report zusammengefasst.

Tipp

Um Ihnen das Abtippen des Codes zu ersparen, können Sie sich das Skript zu diesem Artikel unter folgendem Link auf GitHub herunterladen. Dennoch empfehle ich Ihnen, den Code selbst zu schreiben, um ein Gefühl für die Syntax zu bekommen. einfach.st/r4seocode

Das Google Data Studio (GDS) bietet eine einfache Möglichkeit, Daten aus dem Google-Kosmos – bspw. aus Google Analytics (GA) und der Google Search Console (GSC) – abzufragen, zu visualisieren und als Reports aufzubereiten. Problematisch wird es allerdings, wenn andere Datenquellen angebunden werden sollen. Denkbar sind hier APIs anderer Anbieter wie Sistrix. In solchen Fällen muss der „Umweg“ über Google Spreadsheets und Google Apps Scripts gegangen werden, um die Abfragen selbst zu programmieren. Dies ist durchaus ein valides Vorgehen, das jedoch schnell an seine Grenzen stoßen kann. Denn meist sind Reports nur der erste Schritt. Sobald in diesen Auffälligkeiten erkennbar werden, schließt eine Analyse an, die eine umfassende Beschäftigung mit den zugrunde liegenden Daten erfordert. Derartige Analysen verlangen zumeist komplexe Datentransformationen und -aggregationen, für die Dashboard-Softwares, wie das GSA, mit ihren begrenzten Filtern- und Manipulationsoptionen naturgemäß nicht ausgelegt sind. Ganz abgesehen davon, dass der direkte Zugriff auf die Daten zur Weiterverarbeitung häufig gar nicht gegeben ist. Von großem Vorteil ist es daher, wenn Report- und Analysesoftware in einer Anwendung vereint sind. Diese Anforderungen kann R insofern erfüllen, als die Programmiersprache durch Packages so erweitert werden kann, dass mit ihr sowohl sehr individualisierte Reports erstellt als auch umfangreiche Analysen durchgeführt werden können.

Da Datenanalysen eine sehr komplexe Thematik sind, soll in diesem Artikel jedoch zunächst das Potenzial von R zur Report-Generierung vorgestellt werden. Denn im Grunde sind beide

Aufgaben bis zu einem gewissen Grade sehr ähnlich. Zunächst werden Daten von verschiedenen Quellen abgefragt, dann aufbereitet und schließlich visualisiert. Bei Analysen ist dieser Prozess jedoch iterativ, ohne dass im Vorhinein klar ist, wie das Ergebnis resp. die Erkenntnis genau aussieht. Bei der Generierung von Reports hingegen ist der Ablauf des Datenprozesses vorgegeben. Sie sind daher kleine, standardisierte Analysen, die an einem bestimmten Punkt der Visualisierung aufhören. Das Ergebnisdokument ist somit immer gleich und jederzeit reproduzierbar. Daher eignen sie sich sehr gut, um exemplarisch die Möglichkeiten von R aufzuzeigen, und sind dabei aber gleichzeitig so vereinheitlicht, um durch den Leser selbst nachgebaut zu werden. Ich möchte Sie daher gerne dazu auffordern, das folgende Skript selbst in RStudio umzusetzen.

Hinweis

Der Autor ist sich bewusst, dass gerade für Einsteiger der Anfang etwas holprig sein kann. Aber auch bei erfahreneren Lesern können durchaus Fragen aufkommen. Um hier Hilfestellung zu bieten, die Antworten der Community zugänglich zu machen und den Austausch anzuregen, möchte er Sie ermutigen, in der extra hierfür gegründeten Facebook-Gruppe nachzufragen. Es gibt dabei keine „dummen“ Fragen. Vielmehr hätte der Autor sich in seiner Anfangszeit selbst ein solches Forum mit dem Fokus auf analytische Programmierung und Online-Marketing sehr gewünscht. www.facebook.com/groups/ompyr/

Aber genug der langen Vorrede! In diesem Teil der Serie ist das Ziel, die Sessions aus GA sowie die Performance-Daten aus der GSC abzufragen, um die Packages für diese APIs besser kennenzulernen. Darüber hinaus werden Sie am Beispiel von Sistrix eine eigene API-Abfrage schreiben, um den Sichtbarkeitsindex für eine Domain zu ermitteln. Als Beispiel-Domain dient hier die Website dein-trueffel.de, ein Studentenprojekt an der Hochschule Darmstadt im Modul

„Fortgeschrittene Suchmaschinenoptimierung“, das durch Jens Fauldrath und Stefan Keil betreut wird. Voraussetzung zum Nachvollziehen des Skripts sind eine aktuelle R-Installation (www.r-project.org) und RStudio (www.rstudio.com).

Anlegen eines neuen Projekts in RStudio

Zunächst wird ein neues Projekt angelegt (File → New Project … → New Directory → New Project; Abb. 1). Benennung und Speicherort können Sie frei wählen. Der Einfachheit halber wird für diesen Beitrag das Projekt „report“ (Directory name) unter C:\Users\{USER}\Documents\ (im Eingabedialog durch ~ automatisch abgekürzt; Abb. 2) erzeugt.

Abbildung 1: Anlegen eines neuen Projekts

Abbildung 2: Speicherort und Benennung des Projekts

Das Projekt ist zunächst leer, im File-Explorer finden Sie nur eine RStudio-Projektdatei (report.Rproj), die ignoriert werden kann. Daher wird als erstes über File → New File → R Script (oder den darunter befindlichen Button) eine neue Datei erzeugt und unter der Benennung api_calls.R gespeichert.

Abbildung 3: (1) Neues Skript erstellen, (2) Projekt-Datei und Skript im File-Explorer, (3) Skript-Editor zeigt die erstellte Datei an

Installation und Laden der benötigten Packages

In der soeben erstellen api_calls.R-Datei werden nun die benötigten Packages installiert bzw. – falls bereits vorhanden – aktualisiert und geladen (Abb. 4; 1). Die Google-Packages kennen Sie bereits aus dem Beitrag von Tobias Aubele. Das

Package lubridate erweitert den Funktionsumfang von R für die komfortable Arbeit mit Datums- und Zeitwerten, jsonlite dient zum Abfragen und Verarbeiten von JSON-Dateien. Sobald Funktionen aus diesen Packages verwendet werden, wird explizit darauf hingewiesen. Die vorangestellten Rauten kommentieren den Code aus. Beim ersten Ausführen müssen Sie diese also entfernen. Markieren Sie dann den gesamten Code mit der Maus und führen Sie ihn durch Drücken der STRG- und ENTER-Tasten aus.

Abbildung 4: Installieren bzw. Aktualisieren und Laden der benötigten Packages

In der anschließenden options()-Funktion definieren Sie Ihre Google-Client-ID sowie Ihr -Secret (2). Diese können Sie sich unter console.developers.google.com generieren. Das Argument scopes definiert die Berechtigungen für die GA- bzw. GSC-API. gar_auth() dient schließlich der Authentifizierung (3). Führen Sie den Code das erste Mal aus, öffnet sich Ihr Browser, in dem Sie den Zugriff des Skripts bestätigen müssen. Ist dies erfolgt, finden Sie in Ihrem File-Explorer eine Authentifizierungsdatei (sc.oauth). Bei einem erneuten Ausführen des Skripts verwendet gar_auth() diese Datei, sodass Sie nicht erneut eine manuelle Authentifizierung durchführen müssen – praktisch, denn zukünftig soll das Skript ja automatisch laufen.

Konfiguration wichtiger Variablen

Nun folgt ein Code-Abschnitt, in dem einige Variablen definiert werden (Abb. 5), die an verschiedenen Stellen des Skripts zum Einsatz kommen. Dass die Variablen hier großgeschrieben werden, ist eine Konvention des Autors, die anzeigt, dass die Variablen erst zu einem späteren Zeitpunkt genutzt werden – grundsätzlich könnten Sie die Variablen auch kleinschreiben. Dies ist somit die Stelle im Skript, an der Sie Ihre eigenen Daten eintragen müssen.

Abbildung 5: Definition globaler Variablen, für die die Daten von den APIs abgefragt werden sollen
Im Detail:

(1) Um die passende View-ID Ihrer GA-Datensicht zu ermitteln, können Sie die auskommentierte Funktion ga_account_list() %>% View() verwenden. Der Pipe-Operator %>% übergibt den von der API zurückgegebenen DataFrame mit den Account-Details an die Funktion View(), welche die Tabelle in einem Explorer statt auf der Console öffnet (Abb. 6). Falls der letzte Satz nur ein böhmisches Dorf für Sie war, seien Sie beruhigt, eine ausführlichere Erklärung folgt im nächsten Artikel. Wichtig ist erst einmal nur, dass Sie Ihre View-ID definieren. Entnehmen Sie dazu die gewünschte View-ID aus der Spalte viewId und tragen Sie sie bei der Variable GA_VIEW_ID ein.

(2) Analog dazu verfahren Sie mit der siteUrl für die Variable GSC_PROP.

(3) Die Sistrix-Daten sollen später auf Domain-Ebene abgefragt werden, entsprechend tragen Sie bei SISTRIX_DOMAIN diese ein. Ihren Sistrix-API-Key für die Variable SISTRIX_APY_KEY finden Sie unter app.sistrix.com/account/api.

(4) START_DATE und END_DATE definieren den abzufragenden Berichtszeitraum. In der aktuellen Konfiguration umfasst dieser sechs Monate rückwirkend vom Beginn des aktuellen Monats bis zum heutigen Datum. floor_date(today(), „month“) %m-% months(6) bedient sich dreier Funktionen aus dem Package lubridate. today() gibt das heutige Datum zurück („2018-01-11“). floor_date(„2018-01-11“), „month“) rundet dann zum Beginn des aktuellen Monats („2018-01-01“). %m-% months(6) subtrahiert von diesem Datum sechs Monate („2018-07-01“). %m-% ist ein besonderer Operator aus dem Package, der das Subtrahieren von Monaten unter Berücksichtigung unterschiedlicher Tageszahlen im Monat und Schaltjahren unterstützt. Sie landen mit dieser Funktion also immer am

ersten eines Monats.

Die definierten Variablen können Sie nun im Environment sehen (Abb. 7).

Abbildung 6: Data-Explorer mit den GA-Accounts

Abbildung 7: Environment oder Variable-Explorer listet alle zugewiesenen Variablen

GA-Sessions abfragen

Wenn die Konfiguration abgeschlossen ist, kann mit der ersten API-Abfrage begonnen werden. Als Erstes sollen die Sessions, die aus den Channels Organic Search resp. Direct resultieren, aus GA abgerufen werden (Abb. 8). Um die Abfrage auf diese Dimensionen einzuschränken, werden mittels dim_filter() zwei Filter-Komponenten konfiguriert und in der Variablen channel_filter_organic bzw. channel_filter_direct gespeichert (1). Diese werden an eine weitere Funktion filter_clause_ga4() gegeben, welche intern die Filter in eine für die API verständliche Syntax übersetzt (2). Die Funktion google_analytics() führt schließlich die eigentliche Abfrage aus (3). Hier sehen Sie auch die Variablen für die GA-Datensicht sowie das Start- und Enddatum, die Sie zuvor konfiguriert haben. Führen Sie den Code aus, erscheint in Ihrem Environment die Variable ga_sessions (Abb. 9; 1), die Sie mit der Maus anklicken können und die einen DataFrame mit der API-Antwort enthält (2).

Abbildung 8: API-Abfrage der GA-Sessions

Abbildung 9: DataFrame mit den GA-Sessions

GSC-Performance-Daten abfragen

Ähnlich funktioniert die anschließende Abfrage der GSC-API (Abb. 10). Von dieser werden verschieden Dimensions-Kombinationen abgefragt. (1) Die erste Abfrage ruft die Clicks, Impressions, CTR und Position für den searchType „web“ auf Tagesbasis ab. Auch hier sehen Sie die Verwendung der anfangs definierten Variablen. (2) Dann erfolgt die Abfrage der Metriken für die Kombination aus Datum und Suchanfrage sowie schließlich (3) für die Kombination aus Datum und URL. Auch hier finden Sie nach dem Ausführen der API-Calls die Antwort-Tabellen in Ihrem Environment.

Abbildung 10: API-Abfrage der GSC-Metriken für verschiedene Dimensionskombinationen

Sistrix-Sichtbarkeitsindex abfragen

Zu guter Letzt müssen Sie noch selbst eine API-Abfrage schreiben, denn es gibt natürlich nicht für alle Datenquellen fertige Packages. Sistrix ist so ein Fall, der aber aufgrund der sehr einfachen Abfrage-Syntax ohne großen Aufwand umgesetzt werden kann. Die Abfrage ist nämlich nur eine URL, deren Query-Parameter die gewünschten Abfragewerte aufnehmen. Wird die URL aufgerufen, erhalten Sie als Antwort eine JSON-Datei. Diese URL bauen Sie mit der Funktion paste0() zusammen (Abb. 11; 1). Konkret verbindet die Funktion die in Klammern angegebenen Argumente. Diese bestehen aus den Bestandteilen der Basis-URL für die Abfrage des historischen Sichtbarkeitsindexes (SI), in die die zuvor definierten Variablen für Ihren API-Key sowie die abzufragende Domain eingefügt werden.

Anschließen fragen Sie die zusammengebauten URLs gegen die API an (2). Dazu verwenden Sie die Funktion fromJSON() aus dem Package jsonlite. Die Funktion führt einen GET-Request aus,

nimmt die JSON-Antwort der API entgegen und überführt sie automatisch in eine Liste – ein R-Datentyp zum Speichern hierarchischer Daten. Die Antwort wird zunächst in die jeweiligen Variablen geschrieben. Klicken Sie eine der beiden Variablen im Environment an, sehen Sie, dass das JSON in eine Listen-Struktur überführt wurde (Abb. 12). Hier sehen Sie auch schon den „Weg“, den Sie innerhalb der Liste gehen müssen, um an die SI-Daten heranzukommen. Denn neben diesen erhalten Sie noch weitere Meta-Informationen, wie bspw. die verbrauchten Credits, zurück. Um die Tabelle mit den SI-Daten zu extrahieren, müssen Sie sich an den übergeordneten Listenpunkten entlanghangeln. Dies macht der Ausdruck api_response_desktop$answer$sichtbarkeitsindex[[1]] für den Desktop-SI und überführt den DataFrame in die Variable si_desktop (Abb. 11; 3).

Abbildung 11: Konstruktion der Sistrix-API-Abfrage

Abbildung 12: Liste mit der Antwort der Sistrix-API

Fazit und Ausblick

Damit haben Sie das Ziel dieses Beitrags erreicht. Sie haben ein reproduzierbares Skript geschrieben, dass Sie jederzeit erneut ausführen können. Der Berichtszeitraum der abzufragenden Daten passt sich bei jeder erneuten Ausführung automatisch an. Als Resultat liegen Ihnen die GA- , GSC- und Sistrix-Daten in tabellarischer Form vor. Im nächsten Artikel dieser Reihe wird es dann um die Aufbereitung dieser Daten gehen. Aktuell liegen die Daten auf Tagesbasis vor. Für den Report sollen bspw. die GSC-Daten auf Query- resp. URL-Basis aggregiert werden, um somit Tabellen der Top-10-Suchphrasen bzw. URLs der Vorwoche zu erstellen.