

BACKLINKS, DIE BESUCHER BRINGEN, BLEIBEN WICHTIGER BESTANDTEIL DER SEO

Leonard Metzner

Bekanntermaßen basiert die Erfolgsstory von Google auf der Bewertung von Verlinkungen zwischen Dokumenten für die Berechnung des Rankings. Entsprechend stark bemüht sich Google, Manipulationen an dieser Stelle zu erkennen, um diese wirkungslos zu machen oder sogar zu bestrafen. Dabei zieht die Suchmaschine alle Register: Algorithmus-Updates, manuelle Prüfungen sowie der daraus entstehende und von Mitarbeiteraussagen begleitete Newsflow sorgen beim Thema Backlinks für viel Verunsicherung. Das API Warehouse Leak kann dabei helfen, das Thema Backlinks besser einzuordnen und sich auf die wirklich wichtigen Eigenschaften von Links zu konzentrieren.

Definition von Begriffen

Um voreilige Fehlinterpretationen zu vermeiden, ist es zunächst wichtig, die in den Leaks genutzten englischen Begriffe richtig zu verstehen. Mit der „site“ ist nicht die einzelne Seite (URL) gemeint, sondern das, was wir abseits der rein technischen Bedeutung im deutschsprachigen SEO-Raum oft mit einer Domain meinen (alle Dokumente, URL einer Website). Die „page“ wiederum entspricht einer einzelnen Seite (URL). Unabhängig von diesen im Deutschen oft ungenau verwendeten Begriffen werden in der Dokumentation manche Konzepte anders benannt als in der SEO-Branche üblich. Während bei der SEO von Links die Rede ist, spricht Google in den geleakten Dokumenten auch von „anchors“. Unter dixonjones.com/google-ranking-signals ist eine nach einzelnen Begriffen durchsuchbare Version der geleakten Dokumentation abrufbar. Das Selbststudium ist das beste Mittel, um das, was über das Leak geschrieben wird, auf den Grad der Spekulation zu überprüfen (das gilt auch für diesen Artikel).

eines Links haben können. Interessant wird das vor dem Hintergrund von Links, die aufgrund des Linktexts als Spam markiert wurden (Attribut *phrasesScore* in Modul *IndexingDocJoinerAnchorTrustedInfo*). Folglich ist auch der umgekehrte Fall wahrscheinlich: Es gibt, je nach Inhalt, neutrale (zum Beispiel hier) oder vertrauenswürdige Linktexte (zum Beispiel ein Markenname). Aus diesen lässt sich aber wenig ableiten, weshalb auch auf die Wörter zurückgegriffen wird, die links und rechts vom Ankertext stehen.

Linkquelle und -ziel sollten thematisch zusammenhängen

Über das Attribut *siteRadius* misst Google die Abweichung zwischen *page_embeddings* und *site_embedding*. Anhand von Vektoren könnte also bestimmt werden, inwieweit das Thema einer einzelnen Seite („page“) von dem der gesamten Domain („site“) abweicht. Über *siteFocusScore* wird außerdem erfasst, wie stark eine Domain auf ein Thema fokussiert ist. Für Links könnte das in Zusammenhang mit dem Attribut *anchorMismatchDemotion* heißen, dass die thematische Übereinstimmung zwischen einzelnen miteinander verlinkten Dokumenten wichtig ist. Denkbar wäre darüber hinaus auch die Bewertung des übergeordneten Themas auf Domainebene. Es wäre nur logisch, wenn ein Link dann besonders wertvoll ist, wenn der Inhalt des linkgebenden Dokuments nicht nur mit dem verlinkten Dokument (einzelne Seite) semantisch verwandt ist, sondern mit dem übergeordneten

DER AUTOR



Als Freelancer im Bereich SEO hilft **Leonard Metzner** Unternehmen dabei, mehr relevanten Traffic über die organische Suche zu generieren. Er berät außerdem Agenturen im Bereich SEO, ist Lehrbeauftragter an der THWS und schreibt für verschiedene Blogs und Magazine.

Den Linktext umgebende Begriffe zählen

Im Modul *AnchorsAnchor* bildet Google im Attribut *context2* einen Hash-Wert von Termen in der Nähe des Links. Das bestätigt die schon lange als SEO-Allgemeinwissen bekannte Vermutung, dass nicht nur die Linktexte selbst wichtig sind, sondern auch unmittelbar den Link umgebende Worte einen Einfluss auf die Bewertung

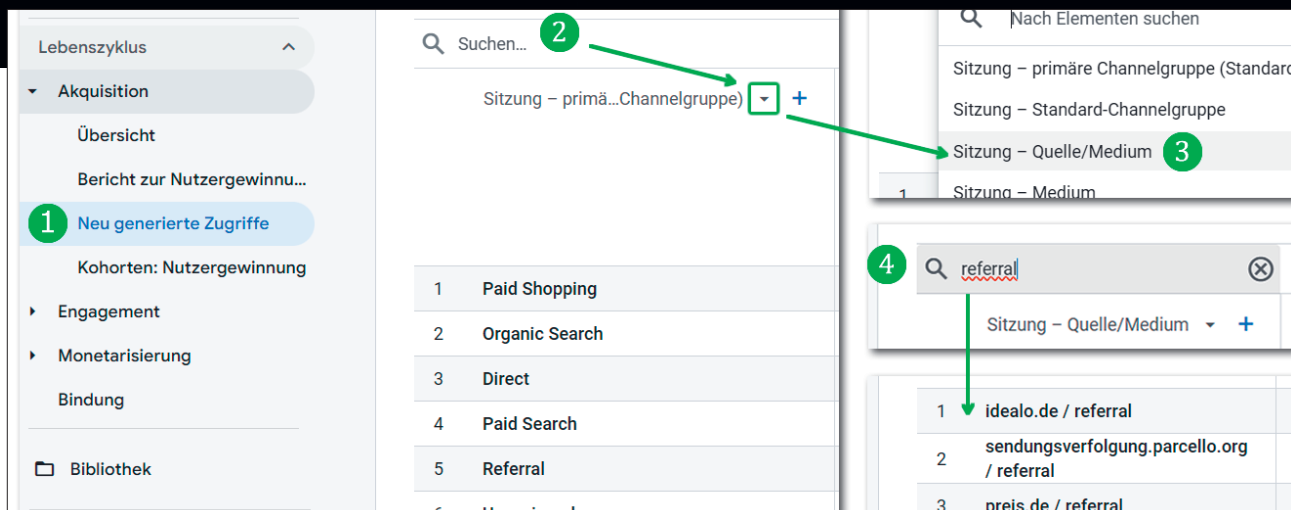


Abb. 1: Welche Links bringen Traffic? Referral Traffic in GA4 anzeigen

Thema der linkempfangenden Domain korrespondiert.

Länderspezifische Backlinks sind wichtig

Über das Attribut *localCountryCodes* werden die Länder erfasst, zu denen das linkgebende Dokument am relevantesten erscheint. Dieser Aspekt ist vor allem für Domains bedeutsam, die verschiedene Länder- und Sprachversionen zur Verfügung stellen. Bei mehreren Länderversionen auf einer Domain profitieren die jeweiligen Verzeichnisse von lokalen Links.

Gewichtet Google bei Links die Schriftgröße?

Im Modul *AnchorAnchor* führt Google das Attribut *fontSize* auf. Eine Beschreibung, die Schlüsse darauf zulässt, wie diese Eigenschaft weiterverwendet wird, existiert nicht. Es wirkt fast etwas zu einfach und aus der Zeit gefallen, Eigenschaften wie die Schriftgröße zu werten. Zumindest als negatives Signal bei einer zu kleinen Schriftgröße wäre das aber durchaus sinnvoll. Denn künstlich erzeugte Links wurden schon immer gerne versteckt. Es kann folglich nicht verkehrt sein, Links deutlich erkennbar zu gestalten, um etwaige Minuspunkte zu vermeiden. Der noch wichtigere Effekt davon ist: Klar erkennbare Links werden häufiger geklickt!

Indexebenen und Klicks

Über das Attribut *sourceType* stellt Google die Qualität einer linkgebenden Seite in Bezug auf deren Indexierungsebene fest. Was sind Indexebenen? Laut Aussagen des Google-Mitarbeiters Gary Illyes können die drei Ebenen RAM, SSD und HDD unterschieden werden. Kurz gesagt werden besonders populäre Inhalte auf besonders schnellen, aber auch teuren Speichern zur Verfügung gestellt. Das ist aber der kleinste Teil des gesamten Index, während die meisten Daten auf günstigeren und langsameren Festplatten liegen. In *sourceType* ist weiter die Rede von Links, die als *TYPE_HIGH_QUALITY*, *TYPE_MEDIUM_QUALITY* oder *TYPE_LOW_QUALITY* markiert werden.

Mit *totalClicks* bezieht Google außerdem die Klicks ein, die ein Dokument erhält. Naheliegender ist somit, die Anzahl an Klicks auch dafür zu benutzen, ein Dokument einer Indexebene zuzuweisen. Der Gedanke, dass Links wertvoller sind, wenn diese aus Dokumenten kommen, die ihrerseits gut gefunden und häufig aufgerufen werden, ist in SEO-Kreisen nicht neu. Die Einführung des Chrome-Browsers und Patente haben schon lange darauf schließen lassen, dass Google Nutzerverhalten in Form von Klicks für die Berechnung des Rankings verwendet. Der entscheidende Satz im Patent US7716225B1 lautet: „*The weights for*

TIPP

Backlinks sind nach wie vor in den meisten Fällen wichtig für ein gutes Ranking in einem dichten Wettbewerbsumfeld. Dass Link nicht gleich Link ist, weiß man bereits seit Langem in der Szene. Aber auch in diesem Bereich haben die Google-Leaks einige aufschlussreiche Erkenntnisse gebracht. Statt wahllos vermeintlich „starke“ Backlinks zu akquirieren oder gar zu kaufen, sollte man differenzierter vorgehen. Gerade in Bezug auf Links von anderen Domains muss man mehr Vorsicht, Sorgfalt und auch Augenmerk walten lassen.

these backward links may be determined based on the user behavior data.“ Ziel sollte es also sein, Links von Seiten zu bekommen, die selbst gut ranken und viel Traffic erhalten.

Vor diesem Hintergrund lässt sich dann auch erklären, weshalb der Einfluss von Backlinks in der Wahrnehmung vieler Suchmaschinenoptimierer in den letzten Jahren abgenommen hat. Gezählt werden vor allem Links, die von Menschen benutzt, also geklickt, werden! Der Shift von der Quantität hin zur Qualität kann als vermeintlich gedämpfte Wirkung von Backlinks an sich fehlinterpretiert werden.

Um bestehende Traffic-Links zu identifizieren, kann, wie in Abbildung 1 gezeigt, innerhalb von Google Analytics 4 der Bericht zur Nutzergewinnung auf „referral“ gefiltert werden. Über die

Sitzung – Quelle/Medium	Seitenpfad und Bildschirmklasse	Name
81 paypal.com / referral	/search	
82 praxistipps.focus.de / referral	/	

/ entspricht der Startseite

Abb. 2: Über den Seitenpfad wird sichtbar, auf welcher URL die Klicks des Traffic-Links landen

zusätzliche Anzeige des Seitenpfads kann ermittelt werden, auf welche URL ein Link zeigt, der Klicks gebracht hat (Abbildung 2).

Die verweisende URL kann dagegen in GA4 nicht immer ermittelt werden, was mit Browsern und Datenschutz zu tun hat. Abbildung 3 zeigt einen Fall, in dem der konkrete Verweis mit einem benutzerdefinierten Bericht angezeigt werden konnte. Für alle anderen Fälle kann bei Google eine Site-Abfrage durchgeführt werden (Abbildung 4). Das funktioniert natürlich nur, wenn die verweisende Seite bei Google indexiert ist und der Markenname/Domainname im Linktext genutzt wird.

Über die Traffic-Ströme zwischen Websites kann Google erkennen, woher Nutzer kommen und wohin sie wollen. Das kann sich im Zeitverlauf natürlich ändern, was auch die Wirkung eines Backlinks ändern könnte. Welche Links in bestimmten Zeiträumen auf Websites, zu denen man keinen Zugang hat, gut funktionieren, zeigen Tools wie Similarweb. Nach der Registrierung sind kostenlose Abfragen möglich, die bereits einige nützliche Daten dazu enthalten.

Abbildung 5 zeigt den ausgehenden Traffic von websiteboosting.com für die Monate März bis Mai 2024. Dabei erhält websiteboosting.com selbst den meisten Traffic organisch. Einige der sichtbarsten Artikel, zum Beispiel „So holen Sie sich Daten von Webseiten: Scraping

Landingpage + Abfragestring	Seitenverweis
278 /c/238	https://www.ecosia.org/
279 /c/238	https://www.gutefrage.net/frage/wo-kann-man-japanische-suessigkeiten-bestellen
280 /c/238	https://www.paypal.com/

Abb. 3: Über die explorative Datenanalyse kann in GA4 der Seitenverweis angegeben werden, der teilweise die vollständige URL enthält

mit XPath & CSSPath“, stammen dabei von Stephan Czysch. Die Websites der Autoren sind in der Autorenbox neben den Artikeln verlinkt. Wer also einen Dofollow-Traffic-Backlink von der Website Boosting möchte, thematisch für Online-Marketer sicher ein Treffer, der sollte viele Artikel schreiben und diese für die spätere Online-Veröffentlichung optimieren.

Autorität und Autoren

Im Attribut *encodedNewsAnchorData* ist von der Klassifizierung von Links von „newsy, high quality sites“ die Rede. Nachdem hier nicht von Seeds gesprochen wird, scheint es dabei nicht um vertrauenswürdige Seed-Sites zu gehen. Interessant ist der Zusammenhang von Newsorientierung und hoher Qualität. Dahinter könnten sich Websites verbergen, die langläufig als Qualitätsmedien bezeichnet werden, aber nicht eindeutig als Seeds markiert wurden, wie zum Beispiel regionale Tageszeitungen. Seeds werden im Leak in der

Beschreibung zu *PagerankNS* ebenfalls erwähnt und sind bereits aus dem Patent US9165040B1 bekannt. Darin ist auch beschrieben, wie ein System Links Längen aufgrund der Eigenschaften der Links und der Eigenschaften der mit den Links verbundenen Seiten zuweist. Daraus berechnet das System dann Ranking-Scores, indem es die kürzeste Entfernung einer Seite zu einer Seed-Seite bestimmt.

Laut Attributbeschreibung wird in *homePageInfo* angegeben, ob es sich bei einer verlinkenden Seite um die Homepage der Website handelt oder nicht. Entspricht die linkgebende URL einer Homepage, können folgende Werte angenommen werden: *NOT_TRUSTED*, *PARTIALLY_TRUSTED*, und *FULLY_TRUSTED*. Links könnten deshalb ein höheres Gewicht bekommen, wenn diese von Startseiten kommen, denen Google vertraut. Der Wert *TYPE_FRESHDOCS* wird dabei als Spezialfall für „freshdoc indexing“ beschrieben und mit *TYPE_HIGH_QUALITY* gleichgesetzt. Das ist

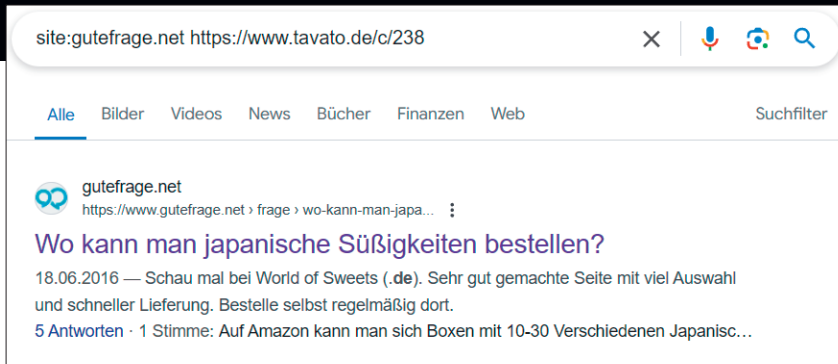


Abb. 4: Je nach Linktext kann mit einer Site-Abfrage die verlinkende URL aufspioniert werden.

ein Anzeichen dafür, dass neuere Links einen stärkeren Einfluss haben könnten, und bekräftigt die Empfehlung von SEO-Experten für Linkbuilding als konstante Aufgabe.

Viele SEO-Tools nutzen für die Bewertung von Domains und deren Linkprofilen eigene Metriken. Eine prominente Ausprägung dieser proprietären Metriken ist die Domain Authority von MOZ, die unter *moz.com/domain-analysis* auch kostenlos abgefragt werden kann. Auf einer Skala von null bis 100 soll das Ranking-Potenzial einer Domain auf Basis des Linkprofils dieser Domain angegeben werden. Während mancher Kunde oder Dienstleister krude Linkbuilding-Vorgaben auf Grundlage solcher Scores entwickelte, belächelten andere diese Zahlen als Marketinginstrument der Toolanbieter. Nicht zuletzt haben schließlich Google-Mitarbeiter auf Rückfragen geantwortet, ihnen sei so etwas wie ein „Website Authority Score“ nicht bekannt. Nun finden wir in den aktuellen Leaks im Modul *CompressedQualitySignals* allerdings die Attribute *siteAuthority* und *authorityPromotion*. Die *authorityPromotion* ist der Beschreibung nach ein von *QualityBoost* *authority.boost* konvertierter Wert, der als Qualitätsbooster fungieren könnte. Ebenfalls scheint ein *unauthoritativeScore* als Qualitätssignal genutzt zu werden, das sich gegenteilig auswirken könnte. Google bestimmt also die Autorität von Websites, wie auch immer diese genau berechnet wird, und nutzt diese vermutlich, um die Ranking-Berechnung zu beeinflussen.

Die Suche nach „Author“ führt in der API-Warehouse-Dokumentation zu vielen Ergebnissen. Über einen Score versucht Google scheinbar, abzuschätzen, wie viel ein Name wert ist, wobei ein Name besser ist, wenn dieser häufiger und in bestimmten Quellen auftaucht („*some sources are considered more authoritative than others*“). Wie viele Gedanken sich Google über Autoren macht, auch wenn das in der Vergangenheit ebenfalls verneint wurde, zeigt auch das Feld *UnmatchedInstitution in ScienceCitation*. Dabei handelt es sich um Autorenzugehörigkeiten, die in einem Dokument gefunden wurden, aber nicht mit bestimmten (bekannten?) Autoren in Verbindung zu bringen waren. Ob es sich bei einer Entität um den Autor des Dokuments handelt, zeigt *isAuthor (true/false)* an.

Die Mischung macht's

Schon seit den ersten linkbezogenen Updates wird in der SEO-Szene über die richtigen Verhältnisse im Linkprofil, zum Beispiel von Nofollow- zu Dofollow-Links oder von Links auf die Startseite zu Deep Links diskutiert. Attribute wie die *spamProbability* zeigen, dass Google Wahrscheinlichkeiten für Spam kalkuliert. In *phraseCount* zählt Google dafür zum Beispiel, wie viele Spam-Phrasen innerhalb der Links von unterschiedlichen (uniquen) Domains gefunden wurden. Laut Beschreibung erfasst Google im Attribut *trustedMatching* zusätzlich die Anzahl an vertrauenswürdigen Links, deren Linktext mit spammigen Termen übereinstimmt. Auch Beispiele

für vertrauenswürdige Quellen können als String im Attribut *trustedExamples* abgerufen werden. Google akzeptiert im Umkehrschluss also eine gewisse Anzahl von Spam-Termen in Linktexten, wenn diese aus vertrauenswürdigen Quellen stammen. In *phraseDays* wird mitgeschnitten, über wie viele Tage hinweg 80 % der Phrasen (vermutlich als Spam einzustufende Phrasen) entdeckt wurden. Spitzen von spammigen Links werden laut *phraseAnchorSpamRate* mit *LINK_SPAM_PHRASE_SPIKE* getaggt. Auch bei Google findet das Paretoprinzip also Anwendung. Wie so oft macht die Dosis das Gift. In einem gewachsenen Linkprofil gehen einige gezielt gesetzt Links also unter, selbst wenn diese als Spam eingestuft werden sollten. Auch wenn sich sicher keine konkreten Zahlen für die richtigen Verhältnisse von bestimmten Linkeigenschaften angeben lassen, ist die Analyse eines Linkprofils somit von größter Bedeutung, bevor Optimierungsmaßnahmen umgesetzt werden. Ein einzelner Link, der als Spam eingestuft wird, führt nicht zu Abstrafungen, kann aber das Fass zum Überlaufen bringen. Das legt auch das im Leak genannte Attribut *badbacklinksPenalized* nahe. Es werden also nicht nur einfach Links entwertet, sondern wegen schlechter Links kann auch ein Penalty vergeben werden.

Die Betrachtung von Zeiträumen, sowohl in Bezug auf die Entstehung von Links als auch bei der Aktualisierung von Content, bestätigt die häufige Annahme, dass SEO Zeit braucht, um ihre Wirkung zu entfalten. Weiterhin unterstreicht dies auch den Ratschlag vieler Suchmaschinenoptimierer, den Aufbau von Backlinks über längere Zeiträume zu strecken und behutsam vorzugehen. Genaue Vorgaben in Form von Werten lassen sich aus den Leaks selbstverständlich nicht ableiten. Sicher ist aber, dass es um das richtige Maß geht, sowohl in Bezug auf absolute Zahlen, Verhältnisse als auch Zeiträume.

Links als verlässlicher Ranking-Faktor

Die Leaks zeigen viele Eigenschaften zur Bewertung von Domains, insbesondere anhand von Links, die Google nutzen kann. In der Vergangenheit wurde viel darüber spekuliert, ob der Algorithmus künftig nicht ohne Links auskommen wird. Offensichtlich, das zeigen die Leaks und auch schlicht die Ergebnisse in der Praxis, ist das bisher nicht der Fall.

Vor dem Hintergrund von KI-Texten und einer damit einhergehenden Flut an semantischem Duplicate Content kann davon ausgegangen werden, dass Links weiterhin eine entscheidende Rolle für das Ranking spielen werden. Ranking-Signale, die Google über Verlinkungen zwischen Websites gewinnt, sind nur schwer manipulierbar, zumindest sehr viel schwerer als alles, was auf der Website selbst passiert. Warum? Die Fähigkeiten von Google, Spam-Links abzuwerten und im Gegenzug hochwertige Links stärker zu gewichten, hat sich stark verbessert. Hochwertige Links, die nicht durchs Raster fallen, können zwar gekauft, aber im Gegensatz zu Texten nicht auf Knopfdruck und fast kostenlos generiert werden. Die Erzeugung von Backlinks erfordert nicht nur den Zugang zur eigenen Website, auf der Content schnell veröffentlicht werden kann, sondern zu anderen Websites beziehungsweise deren Betreibern. Nicht bei jeder Website können Artikel und Links darin plump gekauft werden. Penaltys und linkbezogene Updates haben im letzten Jahrzehnt auf beiden Seiten auch die Qualitätsansprüche an gekaufte Links stark erhöht. Damit gehen stark gestiegene Preise für gekaufte Artikel beziehungsweise Links einher. Man muss also zunächst überhaupt die richtigen Partner finden, genug Budget und Know-how mitbringen, um künstlich Linkaufbau zu betreiben, der von Google nicht direkt ignoriert oder abgestraft wird.

Outgoing Traffic

Top link destinations

Mar 2024 - May 2024 Worldwide Desktop

Domain	Share	Change
 stephan-czysch.de	38.82%	↓ 100.00%
 tech-seo-summit.com	31.66%	↓ 100.00%
 labs.google	19.22%	-
 seo-day.de	10.31%	-

Top referring websites

Mar 2024 - May 2024 Worldwide Desktop

Domain	Share	Change
 ezb.uni-regensburg.de	53.93%	-
 seo-suedwest.de	27.15%	↓ 100.00%
 moodle.lehrerfortbildung-bw.de	18.91%	↓ 100.00%

Abb. 5: Ausgehende und eingehende Links mit dem meisten Traffic für websiteboosting.com laut Similarweb

Gute On-Page-Optimierung und böse Off-Page-Manipulation?

Die immer wiederkehrende Diskussion darüber, ob aktiv Linkbuilding betrieben werden sollte, ist vor dem Hintergrund der dargestellten Bedeutung von Backlinks erstaunlich. Natürlich gibt es „Love-Brands“, die bereits mehr als genug natürlich entstandene Backlinks mitbringen und einem SEO-Manager die Konzentration auf Beratung, Content, Strukturen und Technik erlauben. Doch was ist mit neuen Marken, einer neuen Länderversion, kleinen Firmen, Themen, die weniger attraktiv sind?

Das soll kein Aufruf zum Linkkauf werden, aber ein Denkanstoß in Richtung ganzheitlicher SEO. Ein Rückzug auf reine On-Page-SEO kann mit einem

erlesenen Kundenstamm legitim sein. In vielen Fällen ist es aber genauso falsch wie isoliert durchgeführtes Linkbuilding. Es gibt etwas zwischen dem quantitativ orientierten Kauf thematisch irrelevanter Links für minderwertige Inhalte und der völligen Ignoranz des Faktors Backlinks. Kein Aspekt sollte bei der SEO isoliert betrachtet und optimiert werden.

Je nach Keyword sind Backlinks weiterhin ein wichtiger Anstoß für ein besseres Ranking und bei entsprechendem Wettbewerb die Eintrittskarte für die vorderen Plätze oder den Testlauf von Google für ein Dokument auf der ersten Seite der Suchergebnisse.