



# Duplicate Content

---

Autor: Georg Griesser | Position: Head of IT | Zuletzt aktualisiert: 1. April 2020

Quelle: <https://www.evergreenmedia.at/ratgeber/duplicate-content/>

Er stört vor allem die Suchmaschine und hat das Potential, User gewaltig zu verwirren. Aus SEO-Sicht und teils aus rechtlichen Gründen ist er deshalb nicht empfehlenswert. Die Rede ist von **sogenanntem Duplicate Content**, also doppelten Inhalten, die du wirklich so gar nicht brauchen kannst.

Lies weiter, wenn du unter anderem wissen willst, warum Google Duplicate Content so sehr verabscheut (Spoiler Alert: weil Google User Experience großschreibt und keine Ressourcen verschwenden will), wie du **doppelte Inhalte findest** und wie du sie in Zukunft vermeiden kannst.

---

*The Lowest rating is appropriate if all or almost all of the MC [Main Content] on the page is copied with little or no time, effort, expertise, manual curation, or added value for users. Such pages should be rated Lowest, even if the page assigns credit for the content to another source*

---

Das ist die Aufforderung, die Google für seine [Search Quality Evaluators](#) bereithält. Du siehst: Die Suchmaschine hält nicht allzu viel von duplizierten Inhalten. Schon gar nicht, wenn sie von anderen Seiten zusammengetragen – oder äh, geklaut – sind.

## Was ist Duplicate Content?

Duplicate Content (DC) ist kurz gesagt das **Vorhandensein exakt identischer Inhalte bzw. sehr ähnlicher Inhalte** auf einer einzigen Domain. Auch ein Inhalt, der auf mehreren Webseiten so oder so ähnlich anzutreffen ist, gilt als Duplicate Content. In der Theorie ist es also ganz einfach: Der Begriff meint zwei Inhalte, die zweimal im Internet unter unterschiedlichen URLs zu finden sind und sich entweder entsprechen oder auffällige Ähnlichkeiten aufweisen.

In der Praxis ist die Unterscheidung aber gar nicht immer so einfach. Ja, es kann vorkommen, dass umfangreiche **Textblöcke 1:1 kopiert wurden** und sich ein Konkurrent ganz frech deiner einzigartigen Inhalte angenommen hat.

Genauso gut kann es allerdings sein, dass deine Inhalte **über mehrere Domains, Subdomains oder URLs erreichbar** sind. Auch das ist Duplicate Content, den du erst einmal aufspüren und eliminieren musst. Duplicate Content, der auf anderen Webseiten aufzufinden ist, wird als **externer Duplicate Content** bezeichnet. Alles, was sich innerhalb der eigenen Reihen befindet, ist hingegen **interner Duplicate Content**.



## Wodurch entsteht doppelter Content meistens?

- mit und ohne www möglich
- HTTP und HTTPS möglich
- Trailing Slash
- Facettensuche oder Facettenavigation (Parameter)
- Sitzungs-IDs (Parameter)
- überschneidende Schlagworte
- Paginierung
- alternative Versionen, wie z. B. Druckseiten oder AMP
- Entwicklungsumgebungen
- schlecht umgesetzte Mehrsprachigkeit
- Scraper

## Ab welchem Prozentsatz handelt es sich für Google um Double Content?

Black-Hat-SEO-Quellen behaupten, eine URL gilt für Google als Duplicate Content, wenn diese mindestens 70 Prozent des Inhalts mit einem anderen Artikel oder einer anderen Seite teilt. Hierbei ist jedoch der Kontext wichtig.

Gemeint sind nicht die Inhalte auf deiner Website, sondern **Content für Linkaufbau**. Da Black Hat SEOs [Linkbuilding](#) auf großer Skala mit mehreren [Tiers](#) betreiben, benötigen sie eine gigantische Menge an Inhalten. Sie nutzen dafür [Text-Spinning](#). Ein Duplicate Content Prozentsatz von unter 70 Prozent sorgt ihrer Meinung nach dafür, dass die **URL noch von Google indexiert** und so der **Backlink von dieser URL gezählt** wird.

**Die Inhalte auf deiner Website sollten immer einzigartig sein (Ausnahme: Zitate und Ähnliches).**

## Interner Duplicate Content

Interner Duplicate Content befindet sich **innerhalb deiner eigenen Domain** und tritt besonders gehäuft bei bestimmten Arten von Webseiten auf. Besonders anfällig für Duplicate Content sind [Online-Shops mit Shopsystemen](#) und [Redaktions-CMS](#). Natürlich bedeutet das nicht, dass du mit einem anderen CMS nicht ebenfalls in die Duplicate-Content-Falle tappen kannst.

Schnell passiert ist Duplicate Content, wenn du eine Version deiner Webseite anbietest, die unabhängig von deiner Desktop-Version und [nur auf Mobiltelefone](#) angepasst ist. Sogar **druckerfreundliche Seiten** können schon als doppelter Inhalt gewertet werden.

Auf Seitenebene sind es **vor allem diese Seitentypen**, die unter Umständen betroffen sein könnten. Es schadet also nie, dementsprechend mal aufzuräumen:



- Seitentypen, über die nur bestimmte Inhalte angezeigt werden, d. h. **Tag-Seiten, Filter-Seiten, interne Suchergebnislisten**
- **Kategorienseiten**
- **Produkteinzelseiten**, die in unterschiedliche Kategorien eingeordnet werden können und immer eine eigene URL zugewiesen bekommen (deswegen dürfen Produktseiten niemals die Kategorie in der URL haben)
- Beiträge und Inhalte in **unterschiedlichen Kategorien** (wenn sie die Kategorie in der URL haben)
- **Paginierung vs. Komplettansicht**, wenn nicht alle paginierten Seiten auf die Komplettansicht verweisen

Das sind also die Seiten, die du im Blick behalten musst. Entdeckst du hier Duplicate Content, solltest du daran etwas ändern, selbst wenn das zugegebenermaßen relativ aufwendig sein kann (kann, nicht muss).

Aber Achtung, nicht nur an offensichtlichen Stellen gibt es Duplicate Content. Doppelte Inhalte können sich **extrem gut verstecken**. Das gilt vor allem für Informationen in den **Seitenleisten** und im **Footer**, auch **Boilerplate Content** genannt. Schließlich werden hier einige wichtige Informationen zu deinem Unternehmen angezeigt – und das auf jeder einzelnen URL. Lange Inhalte solltest du hier vermeiden, am besten beschränkst du dich auf **kurze, knappe Informationen**.

|   |  |  |
|---|--|--|
| <b>KONTAKT</b><br><br><b>Evergreen Media AR GmbH</b><br>Fallmerayerstraße 6/3 og<br>6020 Innsbruck, Tirol<br>Österreich<br>E-Mail: <a href="mailto:office@evergreenmedia.at">office@evergreenmedia.at</a><br>Telefon: Terminvereinbarung per E-Mail<br>Preise: ab 3.000 Euro<br>Einzugsgebiete: Deutschland, Österreich<br>(z.B. Wien, Graz, Linz, Salzburg,<br>Innsbruck...), Schweiz | <b>GRUNDLAGEN</b><br><a href="#">Was ist Suchmaschinenoptimierung (SEO)?</a><br><a href="#">Was kostet Suchmaschinenoptimierung (SEO)?</a><br><a href="#">Welche SEO-KPIs sind sinnvoll?</a><br><a href="#">Evergreen Media® SEO-Glossar</a> | <b>UNSERE BELIEBTESTEN INHALTE</b><br><a href="#">Keyword-Recherche: Die Komplettanleitung für deine Keyword-Analyse</a><br><a href="#">WordPress SEO Tutorial: Die Anleitung für eine suchmaschinenoptimierte WordPress-Webseite</a><br><a href="#">Backlinks aufbauen 2019: Der No B.S. Leitfaden für Linkbuilding</a><br><a href="#">SEO für Online-Shops: Die ultimative Anleitung für erfolgreiches E-Commerce SEO</a><br><a href="#">SEO in 2019: Die Top 9 Trends &amp; Entwicklungen im Bereich Google-Optimierung</a> |
| <b>AUS UNSEREM SEO-BLOG</b><br><a href="#">Voice Search &amp; SEO: So verändert Sprachsuche das Suchverhalten</a><br><a href="#">Marketing Automation für Anfänger: So starten Sie die Automatisierung</a>  |  |  |

Wiederkehrende Textblöcke sorgen für „Noise“ (ja, so nennt man das in Black Hat Kreisen). Damit ist gemeint, dass die Relevanz und Qualität der jeweiligen Seite reduziert wird. Deshalb solltest du wiederholende Textfragmente tunlichst vermeiden.

Wenn du deine **Produkte für unterschiedliche Länder** zur Verfügung stellst, musst du ebenfalls ein wenig aufpassen. Während Übersetzungen richtig gekennzeichnet kein Problem darstellen, denkt man normalerweise für deutsche/österreichische/schweizerdeutsche Versionen nicht unbedingt an eine **Kennzeichnung mit „hreflang“**. Das sollte man aber, um Duplicate Content ausschließen zu können.

[WordPress](#) macht vieles gut und richtig. Für viele ist es ein einfaches und intuitives CMS, mit dem schnell eine ästhetisch ansprechende Seite entsteht. Doch aufgepasst: Wenn die gleichen Inhalte sich **komplett auf der Startseite und in einem eigenen Beitrag** finden, ist das nicht optimal. Hier sollte man sich Zeit nehmen (oder einen Experten an Bord holen), um eine passende Lösung zu finden. Zeige in der Blogübersicht immer nur Teaser und nicht den kompletten Content an.



Nicht zuletzt **bei einem Relaunch** sollte auf Duplicate Content ein besonderes Augenmerk gerichtet werden. Ein häufiger Fehler ist nämlich, dass die Testdomain für Außenstehende schon einzusehen und damit auch für den Googlebot zugänglich ist. Ein **Passwort über die „htaccess“** ist deshalb eine wichtige Grundvoraussetzung, um sorgenfrei am Relaunch arbeiten zu können.

## Externer Duplicate Content

Interner Duplicate Content ist nicht gerade ideal, ein größeres Problem stellt aber externer Duplicate Content dar. So wird Duplicate Content bezeichnet, der **über zwei oder mehrere Domains** hinweg besteht. Nicht immer sind diese Copy-and-Paste schuld daran, obwohl Content-Diebstahl und [Content-Scraping](#) bzw. kopierte Inhalte natürlich einen gewissen Anteil am externen Duplicate Content haben.

The screenshot shows a social media post with a profile picture of a person, a 'Teilen' (Share) button, and social media sharing icons for Facebook, Twitter, and LinkedIn. The main text of the post reads: 'Die beste SEO-Agentur finden: Welche Faktoren Sie bei Auswahl lieber außen vor lassen sollten'. Below this, it lists 'Fehler #1: Platzierungen bei Google als Referenz nutzen' and 'Fehler #2: Top 100 SEO-Listen vertrauen'. A highlighted quote states: 'Als Laie ist die Auswahl eines geeigneten Partners unglaublich schwer. Es gibt keine Standards und so gut wie alles ist manipuliert.'

Scrapper-Seiten kopieren deinen Content automatisch. Diesen Content monetarisieren sie dann mit Werbung. Grundsätzlich ist das kein Problem. Wenn du eine sehr kleine und von der Domainautorität schwache Website hast, solltest du neue Inhalte aber immer gleich durch „Abruf wie durch Google“ indizieren lassen.

Doch auch externer Duplicate Content **kann hausgemacht** sein. So zum Beispiel, wenn deine Webseite nicht nur unter einem Domainnamen, sondern gleich **über mehrere Domains** erreichbar ist. In diesem Fall stellst du dir selbst ein Bein, denn Google weiß nicht mehr, was wirklich relevant ist. **Ranking-Probleme** sind die Folge.

Weitere Quellen externer doppelter Inhalte, die für Wirbel im Ranking sorgen können, sind beispielsweise immer dann zu finden, wenn:

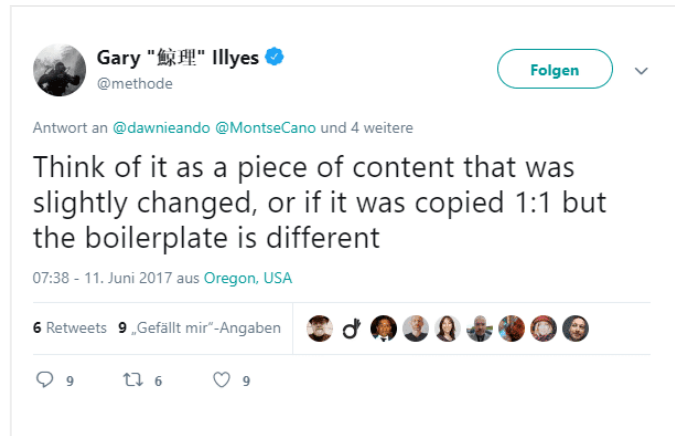
- du die Artikelbeschreibungen einfach **von der Herstellerseite** übernimmst.
- Inhalte **über einen RSS-Feed** eingespielt werden.
- **Pressemeldungen** weiter verbreitet werden.



## Near Duplicate Content

Es steckt schon im Namen: Near Duplicate Content sind doppelte Inhalte, aber eben nicht ganz. Laut Google gibt es zwei mögliche Szenarien, **ab wann Inhalt als Near Duplicate Content eingestuft wird:**

1. Der Inhalt wurde kopiert, aber dabei **leicht abgeändert**.
2. Es handelt sich um eine **1:1-Kopie**, aber der **Boilerplate Content** ist anders.



Googles Gary Illyes sorgt dafür, dass man ein wenig mehr versteht, wie Google die Duplicate-Content-Welt sieht

## Ist alles doppelter Content, was so aussieht?

Nicht alles, was doppelt auftritt, ist Duplicate Content. Und nicht alles, was Duplicate Content ist, macht wirklich Probleme. **Keine Sorgen musst du dir bei diesen Elementen machen:**

- Übersetzungen – immer vorausgesetzt, du hast sie durch „hreflang“ richtig gekennzeichnet
- Zitierte Inhalte – sofern sie im Quelltext richtig gekennzeichnet sind
- Content in Apps

| Duplicate Content   | Duplicate Content   |
|---|---|
| Mit doppelten Inhalten ist gemeint, dass Inhalte gefunden wurden, die über mehrere URLs erreichbar sind. Dieser sog. Duplicate Content sollte unbedingt vermieden werden. Der Inhalt einer Seite darf immer nur über eine eindeutige URL erreichbar sein. Ansonsten stellt man Google vor das Problem, welche der URLs in den Rankings angezeigt werden soll und welche positiven Rankingsignale welcher URL zugeordnet werden. | Duplicate Content means that content is accessible through multiple URLs. This so-called Duplicate Content should be categorically avoided. Each piece of content on a website must only be accessible through one single URL. Otherwise, Google is put on the spot and has to decide which URL to display in the rankings and which positive ranking signals to assign to which URL. |

Dieses Beispiel von [Sistrix](#) zeigt: Nur, weil du Übersetzungen auf deiner Seite hast, heißt das noch lange nicht, dass du doppelten Content am laufenden Meter produzierst

Wenn Google an dieser Stelle die **richtigen Hinweise** erhält, musst du dir hier nicht den Kopf zerbrechen und keine Angst davor haben, dass du im Suchmaschinen-Nirvana verschwindest.



Duplicate Content hat also einigermaßen viele Gesichter. Vielleicht sagt Gary Illyes u. a. deshalb, dass es keine pauschale Antwort auf die Frage gibt, was duplizierte Inhalte eigentlich sind.

## Was Google gegen Duplicate Content hat

Dass die Suchmaschine eine Art Feldzug gegen Duplicate Content angetreten hat, zeigen unzählige Veröffentlichungen zum Thema doppelte Inhalte im Internet. Übrigens geht Google laut eigenen Aussagen nicht automatisch davon aus, dass hinter jedem identischen Textschnipselchen böse Absicht steckt.

Aber das ist nicht Googles Hauptproblem. Das Unternehmen selbst sagt:

---

*Google ist sehr darum bemüht, Seiten mit unterschiedlichen Informationen zu indizieren und anzuzeigen [...].*

---

Das Crawlen und Indexieren von Seiten kostet Ressourcen. Google will diese Prozesse natürlich so effizient wie möglich gestalten, um die Kosten zu reduzieren. Und das ist nicht einfach, bei den unvorstellbaren Mengen an Content, die publiziert werden...



Insgesamt rund 65,02 Millionen Blog-Posts wurden von WordPress-Nutzern im November 2019 veröffentlicht. Gegenüber dem Vormonat entspricht dies einer Abnahme um mehr als fünf Millionen Veröffentlichungen.



Zudem wird durch doppelten Content die URL-Bewertung pro Suchanfrage **erschwert**. Für den Googlebot ist es eine Herausforderung, das Original zu bestimmen und diesem die passenden Signale zuzuordnen.

Und auch für internen Duplicate Content ist **Eindeutigkeit ein wichtiges Gebot**. Ansonsten drohen nämlich verschiedene Seiten innerhalb deiner Domain in einen Konkurrenzkampf um ein und dasselbe Keyword zu treten. Man spricht dann von **Keyword-Kannibalisierung**. Normalerweise meint dieser Begriff, dass mehrere Seiten auf das gleiche Fokus-Keyword optimiert sind. Den gleichen Effekt erzeugst du aber ebenso durch Duplicate Content!

Hier ein Video zu dem Thema: [https://www.youtube.com/watch?v=l4kVABIA9\\_c](https://www.youtube.com/watch?v=l4kVABIA9_c)

Gibt es auf deiner Seite sehr viel Duplicate Content, **zieht der Googlebot irgendwann weiter**, ohne alle wichtigen URLs unter die Lupe genommen zu haben. Der Googlebot hat kein Interesse daran, Googles Index mit doppelten und dreifachen Informationen zu füllen und seine Zeit mit Websites zu verschwenden, die keinen Mehrwert bieten.

Google ist Duplicate Content **nicht** egal, so viel ist klar. Wichtig ist ihnen dabei vor allem die User Experience – und die zu verbessern sollte dein erklärtes Ziel sein.

## Googles Maßnahmen gegen Duplicate Content

Duplizierte Inhalte gilt es zu vermeiden. So weit, so klar. Trotzdem sieht Google noch lange keinen Grund, sofort Maßnahmen zu ergreifen, außer es liegt ein **bewusster Täuschungsfall** vor. In diesem Fall behält Google es sich vor, mit aller Härte vorzugehen und die URL aus der Liste der Suchergebnisse zu entfernen. Doch nicht immer erfolgt zwangsläufig eine Abstrafung durch Google.

Aber wie erkennt Google eigentlich Duplicate Content? Die Suchmaschine crawlt regelmäßig Seiten und identifiziert anhand eines **sogenannten Shingle-Algorithmus** ähnliche oder identische Textelemente. Dazu wird der Text in Schindeln (daher der Name) unterteilt. Sie bestehen immer aus drei einzelnen Wörtern.

Im Anschluss begibt die Suchmaschine sich sozusagen auf Spurensuche. Anhand wichtiger Faktoren versucht sie herauszufinden, welche Inhalte tatsächlich **am relevantesten für die User** sind. Dabei werden folgende Elemente berücksichtigt:

- Alter der Inhalte
- Qualität des Contents
- Relevanz des Inhalts

Problematisch wird das Ganze dann, wenn der Algorithmus nicht deine Original-Seite und deinen Original-Inhalt entdeckt, sondern deine Webseite als weniger relevant eingestuft und diese deshalb herausgefiltert wird.



## Google Panda und Double Content

[Google Panda](#) hat den **Ranking-Algorithmus der Suchmaschine auf den Kopf gestellt**. Insbesondere Webseiten von zweifelhafter Content-Qualität werden seitdem unnachgiebig abgestraft. Seit Februar 2011 (dem Zeitpunkt des großen Panda-Updates, worauf weitere folgten) dürfen bestimmte Probleme, zu denen eben unter anderem Duplicate Content gehört, nicht länger einfach ignoriert werden.

Denn früher betraf Duplicate Content vor allem die entsprechenden Inhalte oder vielleicht die Gesamtanzahl gecrawelter Seiten. Im post-Panda-Zeitalter jedoch kann doppelter Inhalt dazu führen, dass ein Filter für deine Seite angewandt wird, der die **Sichtbarkeit der gesamten Webseite herabstuft**. Im Extremfall sorgt er sogar dafür, dass deine Website nicht mehr indiziert wird. Du tust also gut daran, Duplicate Content nicht als isoliertes Problem zu betrachten, sondern im Gesamtkontext zu sehen.

## Strafe oder keine Strafe, das ist hier die Frage

Während interner Duplicate Content zwar viele Nachteile (Stichwort: Indexierungsprobleme) mit sich bringt, wird er normalerweise von Google doch nicht in diesem Sinne [mit einer Strafe belegt](#). Schwankungen in den SERPs sind aber wahrscheinlich trotzdem dabei.

Mit externem Duplicate Content – insbesondere, wenn es sich wirklich um bewussten Diebstahl von Content anderer handelt – zieht man wiederum tatsächlich den Unmut der Suchmaschine auf sich und **riskiert eine Abstrafung** samt nachhaltigen Schäden. Diese Seiten werden häufig auf die Blacklist verbannt:

- **Scraper Sites**, also Webseiten, die den größten Teil der Inhalte (oft automatisiert) einfach von anderen kopiert haben
- **Content Spinning**, d. h. teils automatisiertes Umschreiben von bereits bestehenden Texten, was nichtsdestotrotz eine Urheberrechtsverletzung darstellt
- [Doorway Pages: Brückenseiten](#) mit Texten, die einzig und allein für die Suchmaschine geschrieben wurden ([mittlerweile fallen für Google auch Local Landing Pages in diese Kategorie](#))

## Urheberschaft erkennen – aber wie?

Best Practice bei Google ist es natürlich, die Seite mit der originalen Urheberschaft zu erkennen. Klar, der Konzern lässt sich hier nur sehr ungern in die Karten blicken. Dennoch ist es wahrscheinlich, dass vor allem darauf geachtet wird, **wessen Content als erstes indiziert** (und damit vermeintlich geschrieben) wurde.

Ein Problem dahinter: Der Googlebot ist auf Seiten mit mehr externen Links häufiger unterwegs.

Dementsprechend kann es unter Umständen passieren, dass die **Fälschung als Original erkannt** wird. Du selbst kannst dazu beitragen, dass von dir geteilte Inhalte erst dann weitergegeben werden, wenn deine Seite tatsächlich indiziert wurde. Das kann sich beispielsweise für Preisvergleichsportale als sinnvoll erweisen.





# Was tun gegen Duplicate Content?

Ansonsten gilt im Kampf gegen duplizierte Inhalte: Stell sicher, dass deine Webseite regelmäßig auf alle SEO-relevanten Fehler geprüft wird. Dazu gehört auch eine regelmäßige Überprüfung auf Duplicate Content. Wenn **technische SEO und Content Hand in Hand** arbeiten, können doppelte Inhalte erfolgreich vermieden werden. Und das ist die wichtigste Strategie, die du gegen doppelte Inhalte fahren kannst.

## Wie kann ich doppelten Content vermeiden? (Schnellübersicht)

- **Informationsstruktur:** Du planst die URL-Struktur deiner Website so, dass Inhalte immer nur unter einer einzigen URL verfügbar sind.
- **Canonical Tags:** Damit kannst du eine Original-URL bestimmen.
- **Meta Robots noindex:** So vermeidest du, dass Google die Inhalte indexiert. Nicht empfohlen!
- **301-Weiterleitungen:** Du kannst URLs mit doppelten Inhalten auf eine einzige URL weiterleiten.
- **Rel="alternate" / Hreflang Tag:** Damit kannst du Google kommunizieren, welche URL für welche Sprachen-Länder-Kombination angezeigt werden soll.

**DUPLICATE CONTENT**

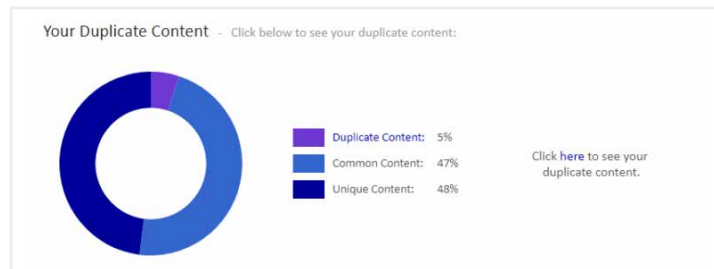
- 1 Mach Gebrauch vom Canonical Tag**  
Das ist vielleicht **der wichtigste Tipp**, den wir dir mitgeben können. Setze rel="canonical", um sicherzustellen, dass **Google die richtige URL indexiert** und rankt. So sucht nicht die Suchmaschine einen Favoriten aus, sondern du.
- 2 Kennzeichne Übersetzungen mit hreflang**  
Wie gesagt, Übersetzungen sind kein Duplicate Content, doch du musst Google vor Augen führen, dass es sich um eine **sprachliche Variante** oder eine andere Sprache handelt. Das **Mittel der Wahl heißt hreflang** und ist im Kampf gegen Duplikation dein Freund.
- 3 Gib dem Googlebot genaue Anweisungen**  
Sag dem Googlebot **klar und deutlich**, ob du möchtest, dass eine Seite indexiert wird oder nicht. **noindex ist der entsprechende Befehl**, den du in deinem Quellcode erteilen solltest.
- 4 Verweh dem Crawler auf keinen Fall den Zugang**  
Es könnte so einfach sein, wenn du einfach in robots.txt dafür sorgst, dass deine Seite nicht mehr gecrawlt wird? Nein, nein und nochmal nein. Das verursacht mehr Probleme, als dass Probleme dadurch gelöst werden. Lieber auf Canonical Tags und Meta Tags setzen.
- 5 Wähle deine bevorzugte Domain**  
Die **Google Search Console** ist eine unglaublich gute Erfindung. Nicht zuletzt, weil du hier die Möglichkeit hast, der Suchmaschine mitzuteilen, was deine bevorzugte Domain ist. Praktisch, wenn du beispielsweise https und http verwendest.
- 6 Kreiere absolut einzigartige Inhalte**  
**Einzigartigkeit** ist das Gegenteil von doppelten Inhalten – eigentlich klar, oder? Nimm dir das zu Herzen und schreib nicht nur bei deiner Konkurrenz ab. Nicht nur Google weiß es zu schätzen, wenn du **phänomenal guten Content** produzierst, auch die User werden gerne auf deiner Seite vorbeischaun.

Um dir einen schnellen Überblick zu verschaffen, wie du Double Content vermeiden kannst, haben wir dir eine kleine Infografik gebastelt.



## Duplicate Content Checker: Tools gegen duplizierte Inhalte

Auf dem Markt gibt es einige Tools, die dir dabei helfen können, duplizierte Inhalte aufzuspüren. Wer sich auf die Suche nach **internem Duplicate Content** begeben möchte, verwendet [Siteliner](#). So erfährst du, zu wie viel Prozent deine Inhalte tatsächlich einzigartig sind und welche Ähnlichkeiten zu anderen internen Seiten es gibt. 250 Seiten können kostenlos untersucht werden.



Wenn du deine URL eingibst, erhältst du einen super Überblick, wie viele duplizierte Inhalte auf deiner Seite zu finden sind. Sieht ganz gut aus bei uns, oder?

Eine ähnliche Funktion erfüllt [Copyscape](#), wobei so auch **externe Quellen** verglichen werden können. Wenn du dir nicht sicher bist, ob eine andere Seite einfach von dir abgeschrieben hat, kannst du mit [Webconfs](#) **zwei Seiten miteinander vergleichen** und hast so das Ergebnis schwarz auf weiß.

**Google Alerts** kann dir – mit den richtigen Voreinstellungen – dabei helfen, externe duplizierte Inhalte zu finden. Wenn du die entsprechenden Textelemente, die du überprüfen lassen möchtest, einträgst (und dabei natürlich deine eigene Seite ausschließt), bekommst du eine Benachrichtigung, wenn der Text erneut im Internet aufgetaucht ist, ohne dass du ihn veröffentlicht hast.

Häufigkeit: Höchstens einmal täglich

Quellen: Automatisch

Sprache: Deutsch

Region: Beliebige Region

Anzahl: Alle Ergebnisse

Senden an: office@evergreenmedia.at

Alert aktualisieren | Optionen ausblenden

Mit dem richtigen Google Alerts Setup wirst du sofort benachrichtigt, wenn jemand deine Inhalte kopiert. Der passende Search Footprint lautet:  
„Ein spezieller Satz aus deinem Artikel.“ -site:deinewebsite.com



## Die tägliche Arbeit eines SEOs

Darüber hinaus kann Duplicate Content am besten vermieden werden, wenn Suchmaschinenoptimierung in Fleisch und Blut übergegangen ist. Aus technischer Sicht bedeutet das, dass du auf eine clevere **URL-Struktur** und die entsprechenden **301-Weiterleitungen** achten solltest.

**Content-Syndication** bedeutet, Inhalte mehrmals verwenden zu können. Damit so kein Duplicate Content entsteht, muss im **Quellcode semantisch** gekennzeichnet sein, woher der Text (das Video, das Bild etc.) ursprünglich stammt. Mit **Canonical Tags** (rel="canonical") kann Google mitgeteilt werden, was die bevorzugte URL ist. Mit **noindex-Tags** stellst du sicher, dass der Googlebot deine Seite gar nicht erst indexiert. Etwas weiter unten erklären wir dir noch mehr dazu.

Nutze die Einstellungen der Google Search Console voll aus, um duplizierte Inhalte zu vermeiden (Crawling > Parameter) und aus dem Index zu entfernen. Achte vor allem darauf, dass die Suchmaschine keine deiner Platzhalterseiten indexiert. Außerdem solltest du dein **Content-Management-System kennen** wie kein zweiter. Nur wenn du weißt, wie Content angezeigt wird, weißt du nämlich auch, wo potenzielle Schwierigkeiten lauern.

Hier ein interessantes Video dazu: [https://www.youtube.com/watch?v=iD-1QVe\\_qvA](https://www.youtube.com/watch?v=iD-1QVe_qvA)

Expertentipp: Bei WordPress sind Unterseiten der Blogstartseite, Tags (Schlagworte), Categories (Kategorien) und Anhangseiten anfällig für doppelte Inhalte. Daher solltest du bei der **Verwendung von Schlagworten und Kategorien** aufpassen und diese so sparsam wie möglich verwenden.

Anhangsseiten hingegen vermeidest du am besten komplett und leitest alle entsprechenden URLs mit Hilfe eines Plugins direkt auf die entsprechenden Bild-URL weiter. Bei Yoast SEO ist diese Funktion standardmäßig bereits aktiviert. Sch au dir ein passendes Video dazu [hier](#) an!

## Strukturen schaffen: Aufräumen gegen Double Content

Im Rahmen der strukturellen **OnPage-Optimierung** solltest du dafür sorgen, dass alle deine Seiten **thematisch absolut eindeutig** sind. **Jedem Inhalt sollte nur eine URL** zugeordnet werden können. Ähnliche Inhalte können auf eine URL zusammengeführt werden. Kreativität (aka Inkonsistenz) bei **internen Verlinkungen** ist denkbar fehl am Platz, genauso wie unnötige Linkvariationen, die unter einer URL zusammengefasst werden sollten.

Wenn auf deiner Webseite landesspezifischer Content veröffentlicht wird, solltest du Google dabei helfen, zu verstehen, dass es sich hier nicht um Duplicate Content handelt. Dazu verwendest du am besten **hreflang Tags** (siehe den Punkt Technische SEO für noch mehr Infos), um verschiedene Sprachen bzw. sprachliche Varietäten zu kennzeichnen.



Außerdem benötigst du **Domains auf oberster Ebene**: mit beispielurl.at gibst du Google schnell und einfach zu verstehen, dass deine Seite Österreich zum Ziel hat.

## Technische SEO: Diese Codes musst du dir merken (& richtig anwenden)

Ohne technische Optimierung ist der Kampf gegen Duplicate Content verloren, bevor er begonnen hat. Einige Hinweise auf die technische Seite des Ganzen haben wir bereits erwähnt, aber weil das Thema so wichtig ist und eine Übersicht sicher nicht schadet, möchten wir ihm ein ganzes Unterkapitel widmen.

Wer sich mit den technischen Seiten der Suchmaschinenoptimierung zu wenig auskennt, für den lautet die Devise: **Wissen vertiefen** (oder natürlich Hilfe von SEO-Experten suchen). Diese Tags sind die Grundlage, um Duplicate Content zu kennzeichnen und damit ‚unschädlich‘ zu machen.

### Meta Robots Tag

Diese Tags enthalten wichtige Informationen über die entsprechende Seite, die für die ‚Augen‘ des Googlebots gedacht sind und entsprechende Anweisungen für den Crawler enthalten. Die Tags sind einigermaßen flexibel, sodass du [verschiedene Befehle für Googles intelligentes Helferlein](#) hinterlegen kannst. Das bevorzugte Format im Header sieht dabei aus wie folgt:

---

```
[meta name="robots" content="noindex,follow"]
```

---

(Achtung: Wir verwenden absichtlich eckige Klammern, statt spitzen Klammern, sonst macht WordPress Faxen)

Was bedeuten die verschiedenen Elemente, die wir am häufigsten verwenden, und warum ist das soeben vorgestellte Format die beste Lösung gegen Duplicate Content, die man sich vorstellen kann?

- **noindex**: Dieser Tag bedeutet genau das, er sagt also dem Crawler, dass er die Website nicht indexieren soll.
- **nofollow**: Auch dieser Tag ist einigermaßen selbsterklärend, er sagt dem Googlebot, dass er Links von dieser Seite nicht folgen soll

Auch eine Kombination verschiedener Befehle ist möglich, wie wir bereits gesehen haben. Würde man rein hypothetisch noindex und nofollow setzen, würde das bedeuten, dass die Webseite weder indexiert wird noch Links von der Seite ausgewertet werden.



Seit Ende 2017 ist bekannt, dass noindex langfristig auch zu nofollow führt, ganz egal ob du die Seite auf die Meta Robots noindex, follow eingestellt hast oder nicht. Mehr dazu erfährst du [hier](#).

**Update vom 11.02.2019:** Google weiß selbst nicht, wie der Algorithmus funktioniert...

**tudhash** 4 points · 5 days ago · edited 3 days ago  
Love this question the67resto! Another related, helpful use case was for **pagination**.

- **Page 1 (e.g. /category/ is index, follow)**
  - I want this crawled + indexed because it's my category's main URL.
- **Page 2+ (e.g. /category?p=2 is NOindex follow)**
  - I want this crawled, but **not indexed** because it only exists to get people and bots to my product detail pages.
- **Product Detail Page (e.g. /p/name-of-product-123 is index, follow):**
  - I want this crawled + indexed since it's the product URL.

**GOAL:** To get Category and Product Detail Pages crawled and indexed AND clearly direct Google not to index no value Page 2+ URLs.

**QUESTION:** What's the recommended way to make sure Page 2+ isn't clogging the index with low value pages aside from proper usage of rel=prev/next?

**garyllyes** 4 points · 2 days ago  
I don't understand the question.  
If the page is **noindex**, it's not "404 eventually", it's just not indexed. If you add "follow" directive to the page, Googlebot will discover and follow the links. That's it, I think

**HalNavel** 2 points · 2 days ago  
Mueller said that links on **noindex** pages will stop being followed after enough time has passed. It surprised us all. With this behavior in mind, should we be making categorization pages like tag archives in WordPress **noindex**? Should we make sure Googlebot has a crawl path of indexable pages to every page we want indexed?

**garyllyes** 5 points · 2 days ago  
Paging [@johnmu](#) I guess.  
Unless we stop visiting the page altogether, which if it's linked well likely won't happen (e.g. it's scheduled because it's linked), we will follow the links

**the67resto** 2 points · 2 days ago  
Thanks for responding, Gary. 🙌  
Hoping [@johnmu](#) can clarify.

Google versteht selbst nicht so ganz, wie Meta Robots noindex funktioniert. John Müller und Gary Ilyes sind sich so gar nicht einig :D

Weitere Informationen zu Robots-Meta-Tags von Google: [https://developers.google.com/search/reference/robots\\_meta\\_tag?hl=de](https://developers.google.com/search/reference/robots_meta_tag?hl=de)

## Canonical Tags

[Diese Tags](#) sind ebenfalls HTML-Elemente im Quellcode. Sie helfen dem Googlebot zu unterscheiden, **was indexiert werden soll und was nicht**. Der Canonical Tag sagt Google, ob es sich um das Original handelt (referenziert sich selbst) oder ob eine andere URL das Original ist. Wer einen Einblick bekommen möchte, was im Moment indexiert wird, sieht sich die Indexabdeckung in der Google Search Console einmal näher an.

---

`[link rel="canonical" href="https://beispiel.com/unterseite.html"]`

---

(Achtung: Wir verwenden absichtlich eckige Klammern, statt spitzen Klammern, sonst macht WordPress Faxen)



Ein Hinweis dafür, dass ein Fehler in der Indexierung vorliegt, ist eine **ungewöhnlich hohe Anzahl** an gecrawlten Seiten, viel mehr nämlich, als es auf deiner Seite wirklich gibt. Grund ist häufig eine URL-Struktur, die nicht konsequent genug umgesetzt wurde oder eben fehlende Canonical Tags für Seiten mit ähnlichem oder gleichem Inhalt.

**Umso wichtiger ist die Kanonisierung deiner URLs**, denn wenn du nicht wählst, wählt der Googlebot für dich. Oder er gibt sogar vorher auf und sieht sich einige deiner wichtigsten Seiten gar nicht weiter an. Was im Moment angezeigt wird, siehst du im [URL-Prüftool](#).

Deine bevorzugte URL gibst du an, indem du den **Canonical Tag manuell setzt**. Für WordPress ist Yoast SEO dein Freund. Und so sieht die Struktur dahinter aus:

Der Nachteil des rel="canonical" ist, dass eine Website dadurch „größer“ werden kann. Auch wenn URLs sich häufig ändern, ist der Tag nicht das richtige Mittel. In diesem Fall sind **301-Redirects sinnvoller**. Sie sind nämlich nicht nur für den Googlebot eine Hilfestellung, sondern auch für User, die so den Weg zur neuen URL finden. Gleichzeitig werden Link Juice und Autorität der alten URL übernommen.

Weitere Informationen zum Canonical Tag von Google: <https://support.google.com/webmasters/answer/139066?hl=de>

## Hreflang

Wir haben es bereits erwähnt: Übersetzungen werden von Google nicht als Duplicate Content bezeichnet und abgestraft. Doch natürlich muss die Suchmaschine irgendwie erkennen, was eigentlich eine Übersetzung bzw. eine Lokalisierung ist. **Das Mittel der Wahl dazu heißt hreflang**. Und so sieht es zum Beispiel aus:

---

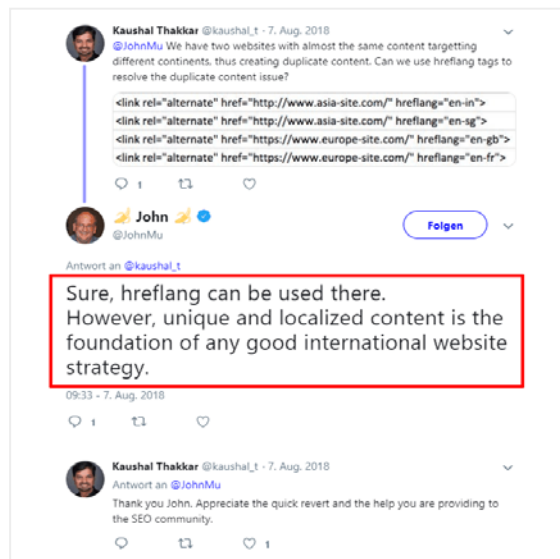
```
[link rel="alternate" href="/beispiel.de/seite.html" hreflang="de-AT" /]
```

---

(Achtung: Wir verwenden absichtlich eckige Klammern, statt spitzen Klammern, sonst macht WordPress Faxen)

[Mit diesem Attribut](#) teilst du Google mit, in welchen Sprachen eine URL verfügbar ist und welche URL zu welcher Kombination aus Sprache und Land ausgeliefert werden soll. Das ist nicht zuletzt für die User Experience wichtig, so findet der User nämlich Inhalte in seiner Sprache bzw. für sein Land sofort bei Google und steigt auf der richtigen Version ein.

Googles John Müller hat auf Anfrage übrigens mitgeteilt, dass **auch bei identischen Inhalten** das hreflang-Attribut dafür sorgt, dass Duplicate Content keine Probleme macht. Gleichzeitig wird allerdings darauf hingewiesen, dass man mit einzigartigem Content für verschiedene Länder (Stichwort: Lokalisierung!) trotzdem die Nase vorne behalten kann.



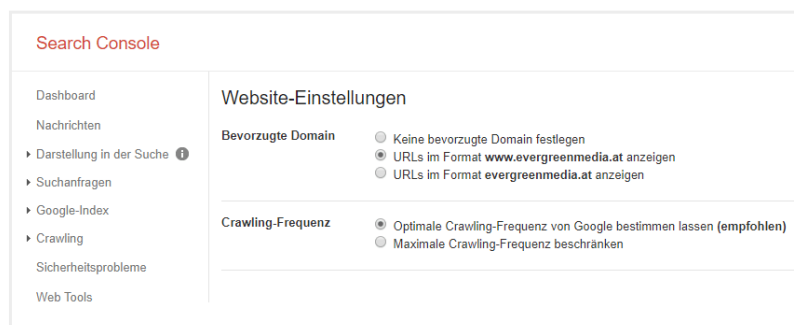
Einzigartige Inhalte kennen keine Ländergrenzen. Am besten überlegst du dir auch, ob deine User in einem anderen Land vielleicht etwas anderes sehen und lesen wollen.

**Expertentipp:** Du hast den Ausbau in weitere Sprachen geplant, bist aber noch nicht soweit? Dann solltest du jetzt schon mit rel="alternate" hreflang="x" planen, damit du deine URL-Struktur nicht eines schönen Tages komplett überarbeiten musst.

Weitere Informationen zum Hreflang Tag von Google: <https://support.google.com/webmasters/answer/189077?hl=de>

## Google Search Console und die bevorzugte Domain

In der Google Search Console kannst du deine bevorzugte Domain festlegen (www vs. ohne www; https vs. http), wichtig ist das, da Google sonst diese zwei unterschiedlichen Formate unter Umständen als unterschiedliche Seiten ansieht.



Durch das Festlegen der bevorzugten Domain gehst du klassischen Duplicate-Content-Problemen schon einmal aus dem Weg. Einfacher wird es nicht :D



Auch [Parameter](#) kannst du in der Google Search Console kategorisieren. So wird deutlich, welchem Zweck sie dienen sollen und weshalb es sich bei ihnen nicht um Duplicate Content handelt. Ein Beispiel, das Google selbst anführt, sind Online Shops. Hier dient beispielsweise der Parameter „country“ als Verweis für das Geotargeting. Es zeigt, dass unterschiedliche Seiten für verschiedene Länder vorgesehen sind. Aber Achtung, Einstellungen an den Parametern sollten **nur dann vorgenommen** werden, wenn man **gewisse Vorkenntnisse** mitbringt und sichergestellt ist, dass Seiten nicht versehentlich vollkommen aus dem Sichtfeld verschwinden.

## Unique statt Duplicate: So schaffst du einzigartige Inhalte

Manchmal scheint es, als würden wir es fast gebetsmühlenartig wiederholen und trotzdem an dieser Stelle noch einmal: **Guter, einzigartiger Content ist die beste Medizin gegen doppelte Inhalte.** Und die Schritte dorthin sind nicht schwierig.

Alles, was zu diesem Zeitpunkt noch nach zu ähnlichem Content aussieht, sollte soweit wie irgend möglich minimiert und eliminiert werden. Auch wiederkehrende Textbausteine gilt es [laut Google selbst](#) möglichst zu vermeiden. Entweder du fasst dich in Sachen Disclaimer am Ende beispielsweise sehr kurz oder aber verlinkst die ausführlichen Zusammenfassungen.

Darüber hinaus lautet das Credo in Sachen Content **so richtig kreativ** zu werden und das eigene Expertenwissen in der großen, weiten Welt des Internets zu verbreiten. Ganz wichtig dabei: Eigenen Content produzieren, der die [Suchintention des Nutzers im Kern trifft](#). Nur zu 80 oder 90 Prozent einzigartig zu sein, reicht nicht aus, um ein Top-Ranking zu erzielen. Unser bester Tipp für Unique Content lautet übrigens, den **Blickwinkel des Users einzunehmen** und so zu schreiben, dass er sich zu 100 Prozent verstanden fühlt.

Schau dir hierzu dieses Video an: <https://www.youtube.com/watch?v=eahyUKEoEuM>

Du willst, dass wir dir genau zeigen, was du an deiner Website ändern musst, um bei Google besser zu ranken?  
[Jetzt SEO-Check buchen!](#)

## Was du gegen Content-Diebstahl machen kannst

Und dann kommt vielleicht doch der düstere Tag, an dem ein Google Alert dich darauf aufmerksam macht, dass deine einzigartigen Inhalte an einer anderen Stelle im Internet aufgetaucht sind. Und das ohne dein Wissen. **Aber was sollst du jetzt tun?**

Das ist wohl abhängig von der Situation selbst. Diese Möglichkeiten hast du:





## Abwarten und Tee trinken

Wenn Google dem Content-Diebstahl bereits auf der Spur ist und noch dazu die andere Seite ein deutlich schlechteres Domain-Rating hat als deine eigene Webseite, musst du dir nicht allzu große Sorgen machen. Zumindest einen [Spamreport](#) wegen einer Verletzung des Urheberrechts solltest du aber trotzdem ausfüllen.

## Mit dem Webseitenbesitzer Kontakt aufnehmen

Oder aber du wendest dich an den Besitzer (oder zumindest Betreuer) der Webseite. Nenne ihm die konkreten Passagen, um die es geht, und weise freundlich, aber bestimmt auf die Urheberrechtsverletzungen hin. Ohne eine **konkrete Frist zur Entfernung** wirst du wahrscheinlich alt und grau werden, ehe dein Content von der anderen Seite verschwindet. Lieber selbst einen Zeitraum angeben.

## Weitere rechtliche Schritte setzen

Es ist die Ultima Ratio, doch wenn der Webseitenbesitzer sich absolut uneinsichtig zeigt, bleibt dir noch immer der Weg, einen **Rechtsanwalt einzuschalten**. Wie gesagt, das ist für alle Beteiligten sehr ungemütlich und sollte deshalb nicht das erste Mittel der Wahl sein. Gleichzeitig ist aber auch klar, dass das Internet eben kein rechtsfreier Raum ist.

Damit dein Content gar nicht erst gestohlen werden kann oder du etwas gegen vermeintliche Inhalts-Diebe in der Hand hast, kannst du übrigens einige Maßnahmen ergreifen:

## Kopierschutz einrichten

So machst du deine Seite unkopierbar, sodass andere viel weniger leicht ‚abschreiben‘ können. Unmöglich ist es allerdings trotzdem nicht und auch die Nutzererfahrung leidet so deutlich. Wir alle wissen von uns selbst, wie häufig wir einfach schnell mal einen Inhalt in die Zwischenablage speichern, um ihn jemandem weiterzuleiten oder später wieder zu finden.

## Hinweis auf Urheberrecht und Copyright

Mit einem solchen Vermerk sind zumindest die rechtlichen Konsequenzen für jeden Besucher deiner Seite klar und deutlich zu erkennen. Im Idealfall wirkt ein prominent platzierter Hinweis abschreckend. Natürlich kann Scraping so nicht verhindert werden, aber zumindest ist dann von vornherein klar: Du verstehst keinen Spaß, wenn es um die unerlaubte Verwendung deines Contents geht.



von moz.com: Wer Angst vor Content-Scraping hat, kann **in den Quellcode eine Art Diebstahlschutz** einbauen. Dabei handelt es sich um einen sogenannten selbst-referenzierenden rel=canonical Link zur bereits bestehenden Seite. Das Attribut verweist dann auf die URL, auf der man sich bereits befindet. Kopieren Content-Diebe den HTML-Code, wird diese Referenz mitkopiert und dein Inhalt wird weiter als Original erkannt.

Keine Panik auf der Titanic. Wenn andere deine Inhalte geklaut haben, sind dir nicht die Hände gebunden – und wer weiß, manchmal motiviert es dich vielleicht sogar dazu, noch bessere und noch aktuellere Inhalte zu veröffentlichen.

## SEO-Fehler, die du bei Duplicate Content vermeiden solltest

Wir wissen es selbst: Eigentlich ist es nicht sinnvoll, am Ende eines Ratgebers alles aufzuzählen, was man so falsch machen kann. Schließlich bleibt im Kopf, was am Ende gesagt wird. Da es aber gleichzeitig eine **gute Zusammenfassung** ist, wagen wir es trotzdem – und hoffen, du setzt unsere praktischen Tipps gerne und erfolgreich um:

### Nicht dem Crawler den Zugriff verwehren

Auf keinen Fall solltest du Duplicate Content in der [robots.txt](#) so kennzeichnen, dass deine Seite nicht mehr gecrawlt wird. Das ist eine nicht sehr elegante Variante, die sich mit rel=“[canonical](#)“ oder [noindex](#) um einiges besser lösen lässt.

### Nicht einfach URLs mit Double Content löschen

Wenn URLs mit Duplicate Content ein Problem sind, weshalb dann nicht einfach die entsprechenden URLs löschen? Die URLs können gelöscht werden, müssen aber vorab auf das Original weitergeleitet werden. Alternativ kannst du sie bestehen lassen und mit dem Canonical Tag das Original bestimmen.

Häufig ist die Ursache für doppelte Inhalte aber eine schlechte URL-Struktur. Nimm dir die Zeit, eine bessere Informationsstruktur auszuarbeiten. Es lohnt sich!



## Nicht mit fremden Federn und Wörtern schmücken

Dein Content sollte zu 100 % einzigartig und authentisch sein. Gute Inhalte, die noch dazu liebevoll aufbereitet sind und für die Besucher deiner Seite einen echten Mehrwert bieten, sind ein echtes Qualitätsmerkmal.

Hier ein Video zu SEO-Texte schreiben: <https://www.youtube.com/watch?v=5cAxsplCmU8>

Duplicate Content unter Kontrolle zu bekommen, ist im Grunde kein Kunststück. Möglich wird es, indem du immer wieder kontrollierst, was bereits auf deiner Webseite passiert ist, sorgfältig arbeitest und technisch immer auf dem neusten Stand bleibst.