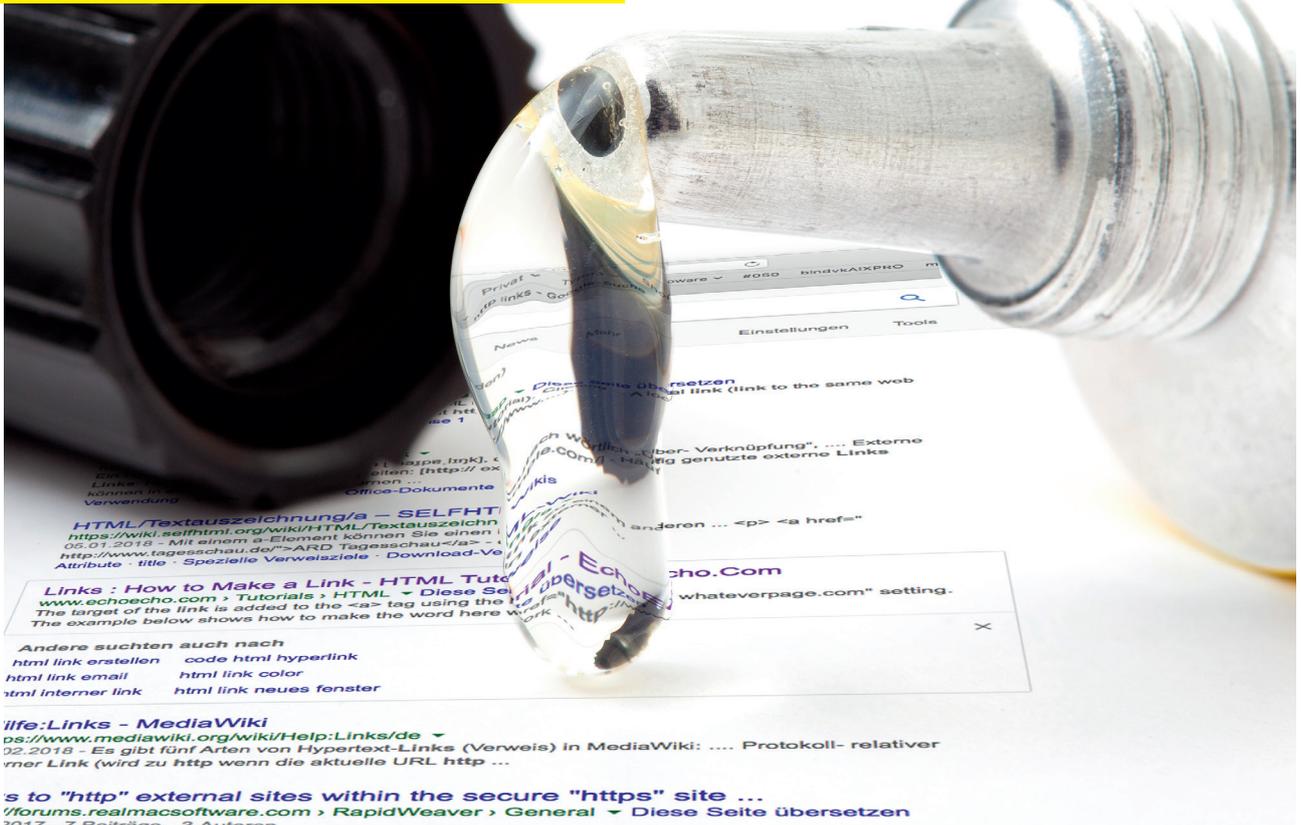


John Müller, Mario Fischer

# RICHTIG VERLINKEN: KLEBSTOFF FÜR WEBSEITEN



Links : How to Make a Link - HTML Tute  
[www.echoecho.com](https://www.echoecho.com) > Tutorials > HTML > Diese Seite übersetzen  
 The target of the link is added to the <a> tag using the  
 The example below shows how to make the word here work

Andere suchten auch nach  
[html link erstellen](#) [code html hyperlink](#)  
[html link email](#) [html link color](#)  
[html interner link](#) [html link neues fenster](#)

ilfe:Links - MediaWiki  
[ps://www.mediawiki.org/wiki/Help:Links/de](https://www.mediawiki.org/wiki/Help:Links/de) >  
 02.2018 - Es gibt fünf Arten von Hypertext-Links (Verweis) in MediaWiki: .... Protokoll- relativer  
 mer Link (wird zu http wenn die aktuelle URL http ...

s to "http" external sites within the secure "https" site ...  
[//forums.realmacsoftware.com](https://forums.realmacsoftware.com) > RapidWeaver > General > Diese Seite übersetzen  
 2017 - 7 Beiträge - 3 Autoren  
 te I have some external links to "http" sites. I wonder if those links will still work after I switch  
 to "https" (sorry for the silly...

Hyperlinks - HTML-Seminar  
[www.html-seminar.de/html-links.htm](http://www.html-seminar.de/html-links.htm) >  
 sind das Salz in der Suppe des WWW. Ohne diese meistens blau unterstrichenen Links  
 WWW nicht funktionieren. Wie, was und wo hier .

ik-Test | Webdienste | WWW | Dienste | URZ | TU Chemnitz  
[u-chemnitz.de/urz/www/link-test.html](http://www.tu-chemnitz.de/urz/www/link-test.html) >  
 shenzentrum: HTML-Link-Test. ... Sie können in HTML-Seiten enthaltene Links  
 er) überprüfen ... (Unterstützt http, https und ftp)

right Labs  
[om/](#) > Diese Seite übersetzen  
 ser. P. R. teaser, Features, Down  
 . Homepage of old Links version

iks

### DER AUTOR



**Johannes Müller** ist seit 2007 ein Webmaster Trends Analyst und Developer Advocate für Search bei Google in Zürich. Er und sein Team verbinden die Welten der Webmaster & Publisher mit den Welten der Entwickler & Produkt-Manager bei Google.

### DER AUTOR



**Mario Fischer** ist Professor für E-Commerce an der Hochschule Würzburg-Schweinfurt und Herausgeber der Website Boosting. Er beschäftigt sich u. a. seit 1996 intensiv mit dem Thema Suchmaschinen.

Führt man das mittlerweile recht komplex gewordene Web auf die Grundbestandteile zurück, bleiben am Ende nur noch zwei übrig: Dokumente bzw. einzelne Webseiten und die Hyperlinks, die sie miteinander verknüpfen. Dokumente gibt es seit Urzeiten, was für das WWW im Kern neu dazukam und eine Nutzung bzw. das „Surfen“ überhaupt erst ermöglichte, waren Links. Während man sich in den Anfängen zwangsweise technisch gesehen mit einfachen Links behelf, ist das Webuniversum heute deutlich anspruchsvoller. Natürlich kann man eine Seite einfach vom Webserver löschen. Aber ist das auch der optimale Weg für Suchende und Suchmaschinen? Wie verlinkt man Seiten in anderen Sprachen für andere Regionen? Wie hilft man Suchmaschinen, Blätternavigtionen oder Breadcrumbs von normalen Links zu unterscheiden? Wie steuert man die Wahrnehmung der Maschine bei Seiten, die mehrfach die gleichen oder fast gleichen Inhalte haben? Die Antwort auf diese und weitere Fragen der internen Verlinkung geben Ihnen John Müller von Google und Mario Fischer.

Während bei der Gestaltung des Contents für einzelne Seiten oft sehr viel Mühe und Sorgfalt aufgewendet wird, unterbleibt das eigentlich fortlaufende Nachjustieren der Verbindungen dieser Seiten oft zu Unrecht. Es ist zwar richtig, dass der PageRank von Google durch sehr viele weitere Signale nicht mehr so dominant wie früher ist, aber nach wie vor berechnet er sich aus den Linkbeziehungen untereinander. Webmaster unterbrechen oder ver- bzw. behindern nicht selten unbewusst den Vererbungsfluss über Links. Die Folge ist, dass beim Ranking unnötig und ohne die Anwendung von „Tricks“ Potenzial verschwendet werden kann, Seiten nur mit zeitlicher Verzögerung in den Index gelangen oder im schlimmsten Fall von Suchenden gar nicht aufzufinden sind.

### Fall 1: Normales Verlinken

Ein Hyperlink kann prinzipiell aus mehreren Elementen bestehen. Nötig sind natürlich das Linkziel und ein sichtbarer Ankertext, auf den per Maus oder per Touch auf sensitiven Bildschirmen geklickt werden kann. Weiterhin kann man steuern, ob beim Klick ein neuer Tab im Browser geöffnet (`target="_blank"`) und ob beim Überstreichen mit dem Mauszeiger ein

erklärender Text in einer gelben Textbox angezeigt wird (`title="Erklärungstext"`). Das Attribut `rel="nofollow"` verhindert, dass Google über diesen Link PageRank bzw. Signale zum Ziel fließen lässt. Dazu gleich mehr. Bei Bedarf kann durch den Klick auf einen Link auch eine hinterlegte Programmlogik mittels JavaScript aufgerufen werden (`onclick="javascript()"`).

In Summe kann ein Hyperlink also z. B. wie folgt aussehen:

```
<a href="http://www.meinedomain.de/" target="_blank"
rel="nofollow" title="Erklärungstext" onclick="javascript();">sichtbarer Linktext</a>
```

Eine besondere Rolle spielt der Ankertext eines Links. Er sagt dem Besucher, was ihn wohl bei einem Klick auf den Link inhaltlich erwarten wird. Insofern sollte man wo immer möglich

aussagekräftige Ankertexte verwenden. Aber auch Suchmaschinen können diesen Ankertext quasi als inhaltliches „Votum“ für die Zielseite verwenden. Wer häufig mit dem gleichen Linktext auf unterschiedliche Seiten verlinkt, darf sich nicht wundern, wenn er damit sowohl die Besucher verwirrt oder verärgert als auch die Suchmaschinen, die maschinell keine eindeutigen Ziele mehr herausrechnen können.

### Fall 2: Auf Ziele mit inhaltlicher Distanz verlinken

Als Besonderheit gilt gemeinhin das „Nofollow“-Attribut. Wird es gesetzt, signalisiert man Google damit, dass man sich vom Linkziel inhaltlich distanzieren bzw. den Link eben nicht setzt, weil die Zielseite der eigenen Meinung nach besonders gute Inhalte aufweist. Verkauft man Links auf andere Webpräsenzen, sollte man den Link tunlichst mit `nofollow` entwerten, um nicht Gefahr zu laufen, eine Abstrafung zu bekommen. Bekanntlich werden Links oft gekauft, um das eigene Ranking durch den damit eingehenden PageRank zu verbessern. Dies ist nach den Richtlinien von Google nicht erlaubt, weil es einen Manipulationsversuch darstellt. Natürlich kann man Links zu anderen Sites verkaufen, damit z. B. dorthin Traffic geleitet wird. Aber um sich von vornherein nicht dem Verdacht eines Verkaufs zur Rankingmanipulation auszusetzen, eben immer mit `nofollow`. Viele Blogsysteme verwenden auch in der Standardeinstellung für

<pre>&lt;a href= "http://www.meinedomain.de/" target="_blank" rel="nofollow" title="Erklärungstext" onclick="javascript();"&gt; sichtbarer Linktext &lt;/a&gt;</pre>	<p><b>BEDEUTUNG</b> (blau= optional)</p> <ul style="list-style-type: none"> <li>-&gt; Linkziel</li> <li>-&gt; öffnet neues Fenster</li> <li>-&gt; Link „entwertet“</li> <li>-&gt; Mousover Effekt</li> <li>-&gt; Prog. Funktionalität</li> <li>» Ankertext</li> </ul>
--	---

Abb.1: Die prinzipielle Anatomie eines Hyperlinks

**NOINDEX/DOFOLLOW –  
FUNKTIONIERT DAS?**

Soll ein Dokument nicht im Index erscheinen, sollte man es im Head mit der Robot-Anweisung „noindex“ kennzeichnen. Was aber passiert mit den Links auf dieser Seite? In der Regel wird diesen keine weitere Beachtung mehr geschenkt. Sie werden weder weiter ausgewertet noch übertragen sie PageRank oder andere Signale. Dies bedeutet, dass mit dem Setzen des Noindex-Tags auch die auf der Seite vorhandenen Links ihre Bedeutung verlieren. Es ist also prinzipiell gesehen unerheblich, ob diese mit „nofollow“ gekennzeichnet sind oder nicht.

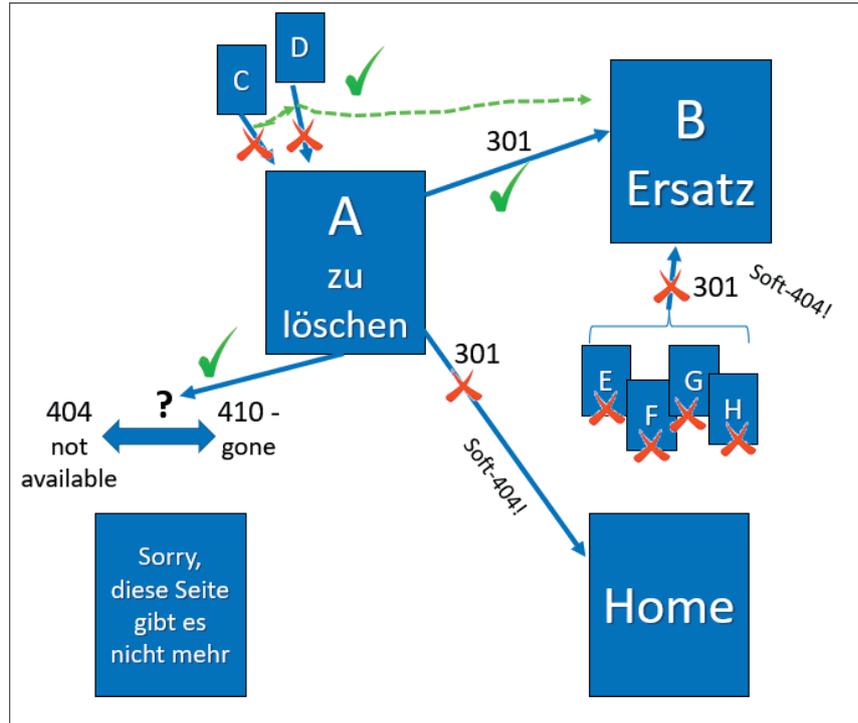


Abb.2: Seiten richtig löschen oder für Ersatz sorgen

Kommentarlinks dieses entwertende Attribut. Nachdem viele Blogs ja von automatischen Systemen mit Spameinträgen und -links geflutet werden, ist dies auch bitter nötig, damit die Blogdomain selbst keinen Schaden nimmt.

Ein Missverständnis besteht oft darin, dass ein Nofollow-Link eingesetzt wird, um das Crawlen der Zielseite zu unterbinden. Das ist allerdings keine gute Vorgehensweise, da die Zielseite auch oft über andere Seiten oder z. B. über die Sitemap verlinkt ist. Sie wird also trotzdem gecrawlt. Einen Ausschluss einer Seite aus dem Index erreicht man tatsächlich nur mit der Robotsanweisung „noindex“ im Head eines Dokuments. Aber auch dies bedeutet nicht, dass die Seite nicht oder nicht mehr vom Bot aufgerufen wird. Sie taucht nur nicht mehr in den Suchergebnissen auf.

In der Szene wird immer wieder kolportiert, dass man mit dem Einsatz von nofollow den PageRank-Flow steuern könnte. Die ursprüngliche Formel zur Berechnung des PageRanks teilt ja die über abgehende Links vererbte Gesamtsumme auf alle Links auf. Hat man 100 abgehende Links, kann jeder rein rechnerisch dann ein Prozent „Wert“ übertragen. Die Idee einiger SEO ist, dass man durch die Entwertung

eines oder mehrerer Links mittels nofollow diesen Wertanteil einsparen könnte und er sich dann auf die restlichen Links verteilt. Jeder Follow-Link hätte dann etwas mehr Power zur Übertragung. Diese Annahme ist allerdings falsch. Der PageRank-Anteil eines Nofollow-Links geht verloren bzw. wird „vernichtet“ und eben nicht anteilig auf die anderen Links verteilt. Diese Art des Versuchs, eine aktive Beeinflussung der Linkberechnung vorzunehmen, wird „PageRank Sculpting“ genannt. Während der Einführung des Nofollow-Attributs funktionierte dies (wahrscheinlich) kurzfristig sogar, aber Google setzte dem Treiben bzw. dieser Art Missbrauch schon vor vielen Jahren ein Ende, indem der Wert wie oben beschrieben einfach gekappt wurde.

**Fall 3: Seiten löschen – weiterleiten oder nicht?**

Beim Betrieb einer Website wird es irgendwann zwangsläufig auch die Notwendigkeit geben, vormals publizierte Inhalte bzw. URLs wieder zu löschen. In einem Shop können das z. B. Produkte sein, die nicht mehr angeboten werden. Hier gibt es verschiedene Möglichkei-

**LINKS IN EINZUKLAPPENDEN SEITENBEREICHEN**

Wie geht Google mit Links um, die in Bereichen liegen, die nicht unmittelbar sichtbar sind bzw. erst nach einer Aktion des Besuchers auftauchen, wie z. B. bei Reitern als Unternavigation auf einer Seite oder sog. Akkordeon-Navigtionen, die man auf- oder zuklicken kann? Bisher gilt, dass sowohl Text als auch Links in zunächst verdeckten Bereichen von Google als nicht unmittelbar wichtigster Inhaltsbestandteil einer Seite gewertet wird. Was wichtig ist, sollte auch immer unmittelbar und ohne weitere Klicks sichtbar werden. Falls das Crawling einer Site aber auf das sog. „Mobile First“ umgestellt wird, genauer, wenn sie künftig nicht mehr mit den Augen des Desktops, sondern des „Smartphone“-Bots gesehen bzw. analysiert wird, ändert sich dies. Dann werden eben auch primär verdeckte Inhalte 1:1 genauso wie sofort Sichtbares gewertet. Dies gilt dann sowohl für die Beurteilung der Desktop- als auch für die der Smartphone-Seiten. Als Sonderfall gelten hier responsive Websites, bei denen es ja nur eine URL für alle Darstellungen gibt. Erfolgt die (individuelle) Umstellung auf den Smartphone-Bot, werden vorher nicht bzw. nicht primär gewertete Inhalte nun voll gewertet.

Statuscode	Kurzbezeichnung	Bedeutung
200	OK	Alles in Ordnung, URL vorhanden und wird/wurde ausgeliefert
301	moved permantently	Umleitung auf eine neue URL, die alte gibt es nicht mehr
302	temporarily moved	Vorübergehende Umleitung auf eine neue URL
404	Not found	Webserver hat unter dieser URL kein Dokument
410	Gone	URL wurde dauerhaft gelöscht, kein Ersatz vorhanden
503	Service Unavailable	Temporärer Serverfehler
5xx	Internal Server Error	Div. interne Serverprobleme

Abb. 3: Kurzübersicht über die wichtigsten Statuscodes

ten, wie man vorgehen kann. Das ersatzlose Löschen einer URL vom Webserver führt dazu, dass beim Aufrufversuch stattdessen eine Fehlerseite (Fehlercode 404) angezeigt wird – sofern der Webserver korrekt konfiguriert wurde. Besucher und Suchmaschinen wissen nun, dass diese Seite mit dieser Adresse nicht mehr existiert. Für Besucher kann das oft frustrierend sein, eine abwehrende Hand vorgehalten zu bekommen. Daher suchen verantwortungsvolle Webmaster in der Regel einen zumindest ähnlichen Ersatz für diesen gelöschten Inhalt. In diesem Fall wird mit dem Statuscode 301 oder 302 auf diese Ersatzseite umgeleitet.

Technisch gesehen bekommt das abfragende System (Browser oder Bot) den Code 301/302 zurück mit der Anforderung, stattdessen die Ersatz-URL aufzurufen. Menschliche Besucher bekommen davon in der Regel gar nichts mit. Der Bot einer Suchmaschine wird sich diesen Status merken und die neue URL an den eigenen Scheduler weiterleiten, der sich baldmöglichst direkt aufruft. Das passiert in der Regel innerhalb von Sekunden. Es sei darauf hingewiesen, dass die ursprüngliche

URL künftig trotzdem immer wieder mal angefordert wird, um zu prüfen, ob die Umleitung noch aktiv ist bzw. die Originalseite doch wieder ans Netz gegangen ist. Das ist normal und kein Grund zu Beunruhigung, falls man solche Aufrufe bei Logfile-Analysen findet. Die Crawlingfrequenz für solche URLs sinkt allerdings deutlich.

Während Code 301 (Abbildung 2) eine permanente Weiterleitung signalisiert, war 302 in den Anfängen des Webs für nur zeitweise Umleitungen reserviert. Diese Fälle, dass man Seiten nur für einen bestimmten Zeitraum umleitet und anschließend doch wieder aktiviert, sind allerdings recht selten geworden. Google behandelt daher beide mittelfristig nahezu als identisch. Einige Systeme wie z. B. Typo3 setzen standardmäßig beim Umleiten einen 302-Code. In SEO-Audits wird dann fast immer bemängelt, dass man besser den Code 301 verwenden sollte. Eine gezielte Umstellung ist allerdings nicht nötig. Eine Weiterleitung mit 301/302 überträgt übrigens alle vererbaren Signale – ohne Dämpfung der Werte.

Möchte man eine ersatzlos gelöschte Seite möglichst schnell aus dem Such-

index bekommen, sollte man statt eines 404-Fehlers besser den korrekten Statuscode 410 (gone) senden (Abbildung 2, Ziffer 4). Dies scheint zwar nur ein marginaler Unterschied zu sein, aber technisch bedeutet ein 404-Fehler, dass die Seite eben gerade nicht verfügbar ist. Die Rückmeldung 410 sagt der Suchmaschine, dass die Seite tatsächlich dauerhaft vom Netz genommen wurde. Sie verschwindet damit schneller aus dem Index. Zudem werden URLs mit einem 404-Code immer wieder mal von Google gecrawlt.

Eine weitere Besonderheit stellen die sog. Soft-404-Fehler dar (Abbildung 2, Ziffer 2). Sie können entstehen, wenn man gelöschte Seiten auf inhaltlich passende Seiten oder viele gelöschte Seiten auf eine einzige Seite weiterleitet. Auch die augenscheinlich einfache Lösung, nämlich gelöschte Seiten einfach immer auf die Startseite zu leiten (Abbildung 2, Ziffer 3), um deren ggf. vorhandenen Qualitätssignale dorthin zu routen, ist also nicht zielführend. Zudem sollte man nicht vergessen, alle internen Links auf die gelöschte Seite auf die neue URL umzuleiten (Abbildung 2, Ziffer 5). Die Weiterleitung per 301/302 funktioniert zwar prinzipiell auch, sodass man sie im Browser beim Klicken nicht bemerkt, aber im Lauf der Zeit entstehen hier möglicherweise lange Weiterleitungsketten, die den Server unnötig belasten. Werden die Ketten zu lang, könnte es am Ende auch mit Google Probleme geben, denn der Bot folgt zunächst nur fünf solcher Kettenelemente in Folge und gibt die weitere Analyse dann an den Scheduler zurück. Der Bot steigt dann an dieser Stelle erst wieder zu einem späteren Zeitpunkt ein.

#### Fall 4: Seiten ohne Links miteinander verbinden

Welchen Sinn kann es machen, Seiten ohne Links miteinander zu verbinden? Eine ganze Menge! Genau genommen verwendet man natürlich

schon Linkziele, aber eben nicht für den Besucher einer Seite sicht- bzw. klickbar. Hier gibt es zwei wichtige Möglichkeiten, die Verlinkung über das Canonical- und das Hreflang-Tag.

### Canonical-Tags

Schätzungen gehen davon aus, dass etwa ein Drittel aller Webseiten sog. Duplicate oder Near Duplicate Content ist. Eine Menge völlig unnötiger Arbeit für Suchmaschinen, die all diese Seiten crawlen müssen, nur um dann am Ende festzustellen, dass die Inhalte so oder fast so schon auf anderen Seiten vorhanden sind. Diese doppelten Inhalte gilt es wegzufiltern, weil man ansonsten Suchenden in den Suchergebnissen im schlimmsten Fall lauter Duplikate vorsetzen würde. Keine schöne Nutzererfahrung.

Nun ist es so, dass sich die Entstehung solcher Duplikate nicht immer gänzlich vermeiden lässt, und sei es nur, weil die Hersteller von Content-Management- oder Shop-Systemen sich darüber keinerlei Gedanken gemacht haben bzw. immer noch nicht machen. Legt man z. B. einen Artikel in einem Shop in mehreren Kategorien an, was durchaus wünschenswert und legitim ist, erscheint ein und derselbe Inhalt oft automatisch in eben diesen unterschiedlichen Kategorieverzeichnissen. Ist ein Teil einer URL anders, wie hier z. B. ein Verzeichnis, ist es prinzipiell eben eine neue Adresse. Je nach System kann das zu einer exzessiven Anzahl an URLs führen, die in keinerlei Verhältnis zu den realen Inhalten mehr steht. Für diese Fälle wurde vor fast zehn Jahren, Anfang 2009, von Google das Canonical-Tag eingeführt. Es wird im Head eines Duplikats bzw. der Duplikate hinterlegt und verweist auf die kanonische Variante, also auf das Originaldokument.

Ein Beispiel:

Der Eintrag im Head einer Seite  
`<link rel="canonical"`

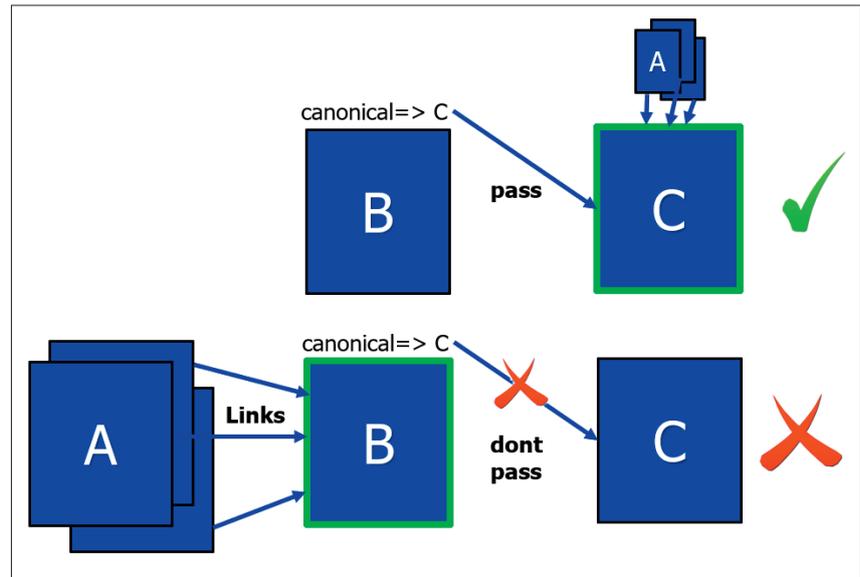


Abb.4: Erhalten Duplikate mehr Links als das Original, kann das Canonical-Tag ignoriert werden

`href="https://www.meinedomain.de/product-1" />`

mit der Adresse *www.meinedomain.de/category-A/product-1* würde Google anzeigen, dass eben diese Seite nur ein Duplikat der Seite *www.meinedomain.de/product-1* ist. Google klappt die Summe der so ausgezeichneten Kopien und das Original zusammen und behandelt sie praktisch wie eine Seite – und im Idealfall erscheint dann im Ranking auch nur diese eine Seite, die kanoni-

sche Variante. Es wird empfohlen, jeder Seite – auch ohne Kopien – immer auch ein selbstreferenzielles canonical im Head mitzugeben. Die Originalseite verweist also direkt auf sich selbst. Stößt der Bot auf eine solche Seite, „weiß“ er sofort, dass es sich hier um das Original handelt und ggf. später auftauchende Kopien können über das URL-Muster schneller identifiziert werden. Das spart Suchmaschinen eine Menge zeitlichen Aufwand und auch Ressourcen, die für die eigene Site dann ggf. für die wirklich wichtigen Seiten zur Verfügung stehen. Kurz gesagt fällt es der Maschine leichter, die wichtige Crawling- und Rechenkapazität zu steuern. So weit die Theorie.

Im Praxiseinsatz kann es allerdings passieren, dass aus Unachtsamkeit intern mehr Links auf ein Duplikat zeigen als auf das Original (Abbildung 4 unten). Nach wie vor werden die Duplikate zusammen mit dem Original als Einheit behandelt, aber hier kann es dann passieren, dass den Suchenden eben nicht mehr das Original (im Beispiel Seite C) in den Suchergebnissen ausgespielt wird, sondern das Duplikat (Seite B), obwohl dieses per Canonical-Tag auf Seite C verweist. Gerade in Shops ist das schnell passiert. Ein Produkt wird (zeitweise) in die Kategorie der Angebote aufgenommen und dann

### VORSICHT BEI SERVERFEHLERN (5XX)

Treten bei Requests am Webserver zu viele 5xx-Fehler auf (Serverfehler), sollte man das genau im Auge behalten und die Ursache(n) suchen und beheben. Bekommt der Googlebot nämlich häufiger solche Fehler als Antwort, reduziert er die Besuchshäufigkeit, weil nicht auszuschließen ist, dass Google selbst den Server durch Abfragen aus- bzw. überlastet. Als Folge davon kommen dann ggf. neue Inhalte erst mit einem gewissen Zeitverzug in den Suchindex. Noch fataler kann sich das bei einem Relaunch auswirken, weil hier sehr viel mehr Seiten in kürzester Zeit gecrawlt werden müssten. Durch die reduzierte Geschwindigkeit kann sich der Zeitraum für einen kompletten Durchgang im schlimmsten Fall vervielfachen.

entsprechend von mehreren Stellen verlinkt. Zunächst ist das kein großer Schaden. Löscht man allerdings die Angebotsseite (aus der eigenen Sicht ja nur eine Kopie), kann es einige Zeit dauern, bis die Suchmaschine wieder „zurück“ zum richtigen Original findet – mit entsprechend zeitlichem Verzug beim Ranking.

Um das Crawlingbudget, also die Anzahl der Seiten, die der Googlebot (individuell) pro Domain aufruft, muss man sich in der Regel übrigens keine Sorgen machen. Der Bot geht prioritätsgesteuert vor und wird in der Regel immer primär wichtige Seiten aufrufen. Erst danach werden z. B. URL mit Parametern, 404-Fehlern oder Seitenkopien erneut besucht, um zu prüfen, ob hier vielleicht aktualisiert werden muss bzw. sich etwas geändert hat.

**Hreflang-Tags**

Eine weitere Besonderheit stellt das Hreflang-Tag (lang steht für language, also Sprache) dar. Auch dieses wird in den Head einer Webseite geschrieben und verweist auf eine inhaltlich gleiche Seite, die aber für eine andere Sprache und/oder ein anderes Land gestaltet wurde. Damit ist es möglich, Seiten die Information über passende regionale Alternativen mitzugeben.

Ein Beispiel:

```
<link rel="alternate" href=
lang="de-ch" href="https://
schweiz.meinedomain.de/seitel.
htm">
```

im Head einer deutschen Seite zeigt an, dass es für Suchende aus der Schweiz mit deutschen Sprachstellungen im Browser eine passende Alternative gibt. Google zeigt zwar die ursprüngliche deutsche Seite in den Suchergebnissen an, aber diese wird bei einem Klick dann ausgetauscht, sodass im Browser die korrekte Seite erscheint. Die erste Kennzeichnung „de“ steht dabei immer für die Sprache, die zweite (ch) für das Land. „fr-ch“ würde

also beispielsweise für die Schweiz ausgespielt werden, und zwar für Browser, die Französisch als Sprache hinterlegt haben. Es empfiehlt sich auf jeden Fall, immer auch eine Rückfallvariante zu hinterlegen, falls Zuordnungen, aus welchen Gründen auch immer, nicht funktionieren sollten: die Variante x-default. Sie führt auf eine URL, die am besten für alle Anfragen passt, z. B.:

```
<link rel="alternate"
href="https://meinedomain.de/"
hreflang="x-default" />
```

Natürlich sollten alle Verweise auf Gegenseitigkeit beruhen. Das bedeutet, wenn eine de-de-Variante z. B. auf fr-ch verweist, muss diese Seite einen Rückverweis auf de-de im Head tragen. Dies gilt für alle Varianten, die man verwendet. Wird ein solcher Verweis „gebrochen“, d. h., es fehlt der entsprechende Rückverweis, wird (nur) diese Länder-/Sprach-Verknüpfung ignoriert. Alle anderen Ketten bleiben aber intakt. Insofern sind die manchmal dramatisch klingenden Rückmeldungen einiger Tools, man liefe Gefahr, sämtliche Verbindungen zu verlieren, tatsächlich unbegründet.

**Fall 5: Besondere Links kennzeichnen**

Auch für Fälle, wo Links nicht die übliche Verweisbedeutung haben, sollte man Vorsorge tragen. So haben z. B. Breadcrumb-Links eine etwas andere Bedeutung, weil sie eine Hierarchie ausweisen. Auch Blätterlinks in eigenen Suchergebnisseiten oder Produktlistings, die nicht mehr auf eine Seite passen, wollen eigentlich „nur“ die nächste Seite erreichbar machen und tragen keinen inhaltlichen Sinn. Es ist keine schlechte Idee, solche Links für Google und andere Suchmaschinen besonders zu kennzeichnen, damit sie entsprechend erkannt und ggf. anders behandelt werden können.

**Breadcrumbs**

Idealerweise gibt man einem Breadcrumb-Link den Title, die URL und ggf. die Information „child“ maschinenlesbar mit auf dem Weg. Für eine Breadcrumb wie z. B.

Kleider > Minikleider > schwarze Minikleider

sähe die entsprechende Metainformation bzw. die Programmierung beispielsweise so aus:

```
<div itemscope itemty-
pe="https://data-vocabulary.org/
Breadcrumb">
```

```
<a href="http://www.ihresite.
de/kleider" itemprop="url">
<span itemprop="title">Klei-
der</span> </a> > </div>
```

```
<div itemscope itemty-
pe="https://data-vocabulary.org/
Breadcrumb" itemprop="child">
```

```
<a href="http://www.ihresite.
de/kleider/mini" itemprop="url">
<span itemprop="title">Minik-
leider</span> </a> > </div>
```

```
<div itemscope itemty-
pe="https://data-vocabulary.org/
Breadcrumb" itemprop="child">
```

```
<a href="http://www.ihresite.
de/kleider/mini/schwarz" item-
prop="url">
<span itemprop="title">
schwarze Minikleider</span> </a>
</div>
```

Die Algorithmen können auf diese Weise erkennen, dass es sich nicht einfach nur um drei Links handelt, sondern dass diese eine Hierarchie darstellen. Das erleichtert die maschinelle Aufarbeitung der Sitestruktur. Das Mark-up kann dabei als JSON-LD, Microdata oder RDFa hinterlegt werden. Weiterführende Infos und Beispiele dazu findet man bei Google unter <http://einfach.st/breadcrumb>.

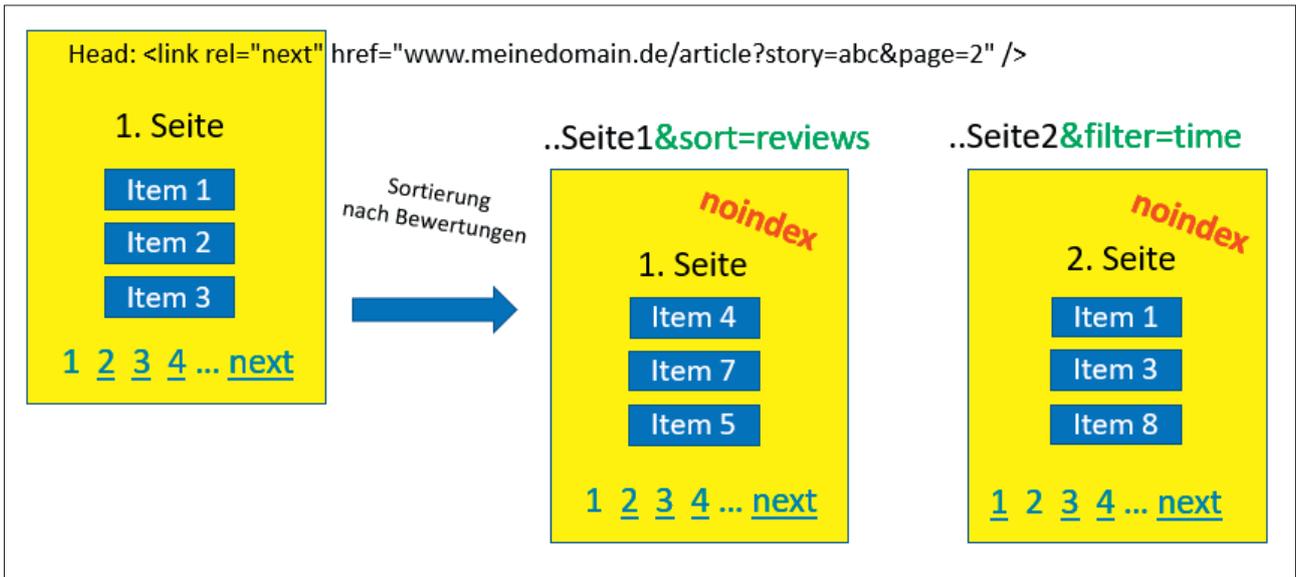


Abb. 5: Normale Paginierungen müssen kein noindex tragen; Ausnahmen sind Seiten mit angehängten Sortier- und Filterparametern

**Paginierung bzw. Blätterlinks: rel="prev" und rel="next"**

Seiten, durch die man weiterblättert, können Suchmaschinen durchaus verwirren bzw. in Verbindung mit Such- und/oder Filterparametern zum sog. unendlichen Crawl führen. Daher ist sehr wichtig, auch diese Art der Verlinkung gesondert zu programmieren.

Dazu wird im Head der ersten Blätterseite (www.meinedomain.de/article?story=abc&page=1) die folgende Information hinterlegt:

```
<link rel="next" href="https://www.meinedomain.de/article?story=abc&page=2" />
```

Damit wird maschinenlesbar ausgewiesen, dass es eine nachfolgende Seite (page=2) gibt. Auf der zweiten Seite (www.xy.de/article?story=abc&page=2) wird dann die Vorgänger- und die Nachfolgerseite im Head angegeben:

```
<link rel="prev" href="https://www.meinedomain.de/article?story=abc&page=1" />
```

```
<link rel="next" href="https://www.meinedomain.de/article?story=abc&page=3" />
```

Und so setzt man dies auf allen folgenden Seiten um. Es wird jeweils mit „prev“ auf die vorhergehende Seite verwiesen, mit „next“ auf die nachfolgende. Auf der letzten Seite der Blät-



Abb. 6: Trotz Blockierung per robots.txt können Seiten bei Google auftauchen

Methode	Linkjuice wird vererbt	Seite wird gecrawlt	Seite wird indexiert
Blockierung in der robots.txt	Nein	Nein	Jein (1)
Meta Tag "noindex"	Ja - aber nur temporär!	Ja (2) - aber weniger häufig!	Nein
Meta Tag "nofollow"	Nein	Ja	Ja

(1) Bei eingehenden Links kann die Seite in den Ergebnissen auftauchen  
 (2) Nur wenn andere, follow-Links auf die Seite zeigen

Abb. 7: Unterschiedliche Methoden der Blockierung führen zu unterschiedlichen Ergebnissen

ternavigation kann man natürlich dann nur noch auf den Vorgänger verweisen, weil ja kein Nachfolger mehr existiert.

In der Branche wird oft leidenschaftlich diskutiert, ob man hier auf den Folgeseiten mit „noindex“ oder dem canonical arbeiten sollte, damit man sie aus den Suchergebnissen heraushalten kann. Das ist allerdings gar nicht nötig, wenn „prev“ und „next“ korrekt eingesetzt werden. Google erkennt die Paginierung und klappt auch diese Seiten automatisch zusammen – im Idealfall sollte dann auch

hier jeweils nur die erste Seite in Suchergebnissen auftauchen. Im Gegenteil wäre die Verwendung des Canonical-Tags sogar streng genommen ein Art Missbrauch, denn die Inhalte sind ja eben meist nicht (fast) identisch.

Verwendet man weitere Parameter, die an die URL angehängt werden, um z. B. Filter oder Sortiereinstellungen auszuspielen, sollte man diese Versionen im Head mit „noindex“ auszeichnen (Abbildung 5). Durch Mustererkennung kann Google in den meisten Fällen die Art und Struktur der

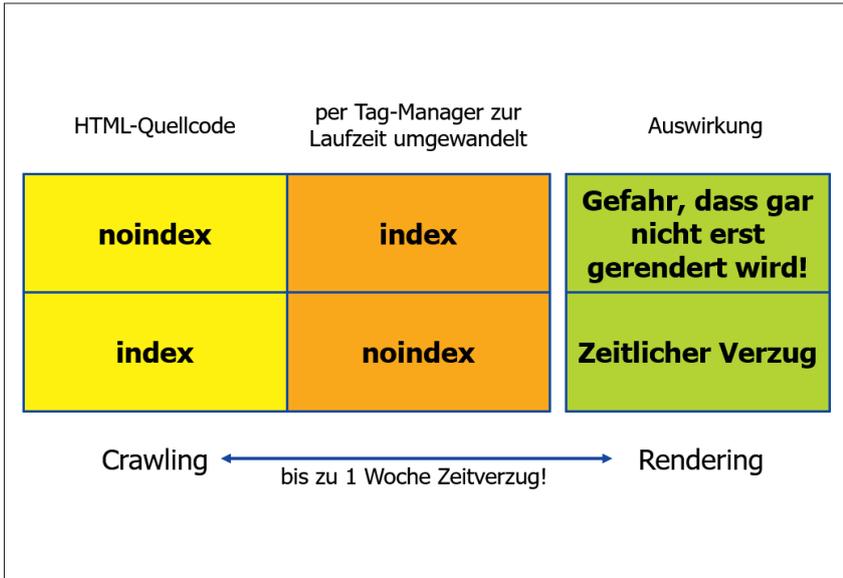


Abb. 8: Mit essenziellen Tags über den Tag-Manager herumzuspielen, ist meist keine gute Idee

Parameter separieren und diese Seiten bzw. URLs, die ja über Funktionslinks („Sortieren nach ...“) immer wieder auftauchen, dann ignorieren. In der Searchconsole findet man oftmals auch solche automatisch erkannten und hinterlegten Anhangfragmente. Die Verwendung des Canonical-Tags empfiehlt sich für Paginierungen nicht.

### Fall 6: Einzelne Seiten blockieren

Möchte man das Auftauchen bestimmter Seiten im Suchindex von Google vermeiden, gibt es eigentlich nur eine vernünftige Methode, nämlich das Auszeichnen im Head mit noindex. Eine Blockierung über die robots.txt verhindert das Auffinden nämlich nicht. Abbildung 6 zeigt dies beispielhaft. Tatsächlich crawlt der Googlebot die „verbotene“ URL zwar nicht und kennt folgerichtig auch nicht deren Inhalt. Aber wenn genügend externe Signale z. B. in Form von Links darauf vorhanden sind, wird die URL (mit einem entsprechenden Hinweis) trotzdem ausgespielt.

Wie schon weiter oben erwähnt, ist die Kennzeichnung von Links mit nofollow auch kein probates Mittel, um ein Crawlen und auftauchen im Index zuverlässig zu verhindern (Abbildung 7).

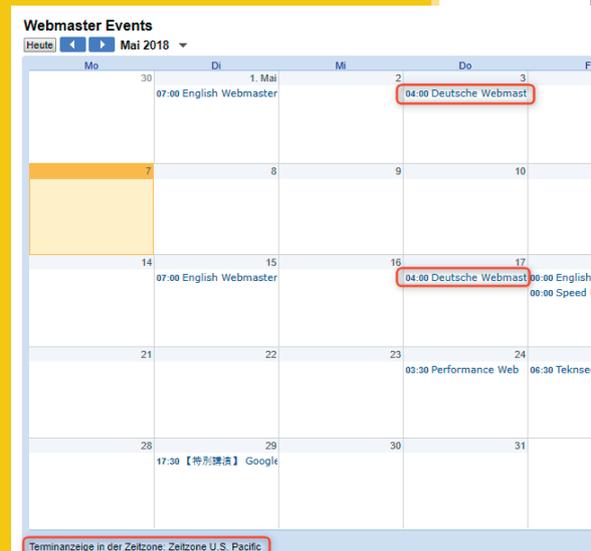
Fatal kann sich folgende Vorgehensweise auswirken, auf die man ggf. durch eine plötzliche Abmahnung und eine nachvollziehbare Panik verfällt: Man löscht die Seite und verbietet das weitere Crawlen per robots.txt. Da Google die Seite nicht mehr aufrufen darf, bleibt die letzte Variante zunächst für einen gewissen Zeitraum in Cache bestehen und kann von dort aus im Suchergebnis auch aufgerufen werden – oft sehr zur Freude des abmahnenden Rechtsanwalts. Ebenso ungünstig ist es, die Seite mit noindex zu versehen und das Crawlen über

die robots.txt zu untersagen. Der Bot erhält dann nämlich keine Kenntnis mehr von der Änderung (noindex!) und lässt ebenfalls die alte Variante für einen gewissen Zeitraum stehen. Den besten, weil schnellsten Effekt erzielt man dadurch, dass man die Seite löscht (Servercode 410). Soll sie zwar weiterhin bestehen bleiben, aber nicht mehr im Index auftauchen, setzt man sie zunächst im Head auf noindex und ruft die URL dann manuell über die Searchconsole ab (Menüpunkt „Crawling“ und dann „Abruf wie durch Google“). Durch diesen Vorgang bekommt Google am schnellsten Kenntnis vom Indexierungsverbot. Dies kann man nachprüfen, wenn man sich den gerenderten Quellcode der Seite in der Searchconsole anschaut. Dort sollte und muss dann die neue Anweisung noindex mit enthalten sein.

Eine Anweisung robots.txt „disallow“ kann auch auf Seiten Sinn machen, auf denen User (komplexere) Berechnungen durchführen können, die nennenswerte Serverkapazität benötigen. Damit nicht auch jeder Botbesuch solche Berechnungen auslöst, blockiert man die Seite sinnigerweise einfach per robots.txt.

## JOHN MÜLLERS WEBMASTERSPRECHSTUNDEN

Falls Sie auf dem Laufenden bleiben wollen oder um vielleicht selbst direkt Fragen an John Müller stellen zu können, empfiehlt sich die Webmastersprechstunde, die in unregelmäßigen Abständen via Hangout stattfindet. Kontaktmöglichkeiten finden Sie unter [www.google.com/webmasters/connect](http://www.google.com/webmasters/connect). Dort finden Sie auch weiter unten einen aktuellen Kalender, in dem die Livetermine angekündigt werden. Bei Bedarf können Sie diesen Googlekalender sogar zu ihrem eigenen hinzufügen und die Termine dort nach Bedarf ein- oder ausblenden. Achtung, die Zeiten dort sind in der USA Pacific Time angegeben. Für Deutschland sind jeweils acht Std. zu addieren, plus eine weitere Stunde in der Sommerzeit.



### Vorsicht bei Codeänderung per Tag-Manager

Der Google-Tag-Manager ist ein machtvolles Instrument. Und nicht erst seit der Star-Wars-Saga weiß man, dass man mit der Macht sorgsam umgehen sollte. Prinzipiell kann man damit zur Laufzeit, also beim Aufrufen einer Website, Inhalte fast beliebig verändern. Schnell sind SEO daher auf die Idee gekommen, die Indexierung bequem steuern zu können, ohne den Quellcode selbst verändern zu müssen. Dies ist allerdings keine gute Idee. Setzt man z. B. (bei neuen Seiten) standardmäßig alles auf noindex und ändert das bei Bedarf per Tag-Manager auf index um, wird Google diese Änderung gar nicht mehr mitbekommen. Stößt nämlich der Bot auf die Noindex-Anweisung, hat das zur Folge, dass die Seite gar nicht

erst zum Rendern weitergereicht wird. Sie ist ja unwichtig und soll nicht in den Index. Aus Sicht von Google macht es also keinen Sinn, Kapazität für das Rendern so einer Seite aufzuwenden – wo ja dann erst erkannt werden könnte, dass die Anweisung sich jetzt dynamisch verändert hat (Abbildung 8). Der umgekehrte Fall, zur Laufzeit per Tag-Manager die Anweisung noindex einzusetzen, bedeutet zumindest einen gewissen zeitlichen Verzug, bis Google dies tatsächlich erkennen kann. Zwischen Crawling und Rendering kann unter ungünstigen Umständen nämlich bis zu einer Woche Zeit liegen.

### Fazit

Die richtige Verlinkung bzw. Verbindung zwischen Webseiten ist tatsächlich kein Hexenwerk. Wenn man einige

Besonderheiten beachtet, wird man mit häufigeren Botbesuchen (Aktualität) belohnt und durch das Herausrechnen spezieller Links wie z. B. Paginierungen oder das richtige Zusammenlegen von Duplicate Content erhält man qualitativ hochwertige, weil „richtigere“ Suchergebnisse. Zudem kann bei richtiger Verlinkung am Ende auch das Ranking profitieren, der Besucher und damit auch der Websitebetreiber profitieren sowieso. Es lohnt sich in aller Regel immer, dafür zu sorgen, dass auch die Algorithmen von Google und anderen Suchmaschinen „besondere“ Linkarten und -strukturen besser verstehen. ¶

### „Glückwunsch!!!

- innovativ - hochwertig - persönlich - effizient -  
beschreibt das Magazin und die Zusammenarbeit mit dem Team der Steuerkanzlei Kempin.

Vielen Lieben Dank!  
(freuen uns auf die nächsten 50 ... ☺ )“

### Steuerkanzlei Kempin



Sofia Martina Kempin

Steuerberaterin - Dipl. Betriebswirtin (FH)

Fachberaterin für Internationales Steuerrecht

[www.steuerkanzlei-kempin.de](http://www.steuerkanzlei-kempin.de)

