

Axel Scheuring

»Canonical-Tag – das gefährlichste Tag, seit es HTML gibt

Das Canonical-Tag ist eine gern genutzte Möglichkeit, um Duplicate Content (doppelte Inhalte) zu beseitigen. Es funktioniert sehr effektiv, was zwar viele Vor-, aber leider auch einige Nachteile hat, denn durch genau diese Effektivität werden Fehler umgehend und sehr hart „bestraft“. Im schlimmsten Fall fällt eine Website komplett aus den Suchergebnissen. Deshalb sollten Sie wissen, worauf Sie achten müssen, wenn Sie das gefährlichste Tag, seit es HTML gibt, einsetzen. Axel Scheuring gibt dazu wertvolle Tipps.

Im Prinzip ist das Canonical-Tag eine tolle Sache. Seit Februar 2009 leistet es vielen Website-Betreibern wertvolle Dienste im Kampf gegen [Duplicate Content*](#). Damals verkündeten alle bedeutenden Suchmaschinen, dass man mithilfe des Canonical-Tags Duplicate Content entfernen kann. Und das Schöne dabei: Es funktioniert! Zumindest bei Google. Und da Google einen Marktanteil von circa 90 Prozent in Deutschland hat, erscheint es weniger tragisch, dass man Duplicate Content mit dem Canonical-Tag bei Bing und Yahoo nicht ganz so perfekt entfernen kann. Im Großen und Ganzen funktioniert es also.

Doch was genau ist eigentlich das Canonical-Tag?

Das Canonical-Tag sieht wie folgt aus:
`<link rel="canonical" href="http://www.ein-elektronikshop.de/apple/ipads" />`

Spricht: Es wird über eine Codezeile im <Head> eines HTML-Dokuments platziert. Diese Codezeile bewirkt, dass man mit der Webadresse (URL), die man im HREF-Attribut angibt, die Originalversion einer Webseite benennt. Das ist praktisch, wenn eine Webseite unter mehreren URLs erreichbar ist. Denn hier entstünde sonst Duplicate Content. Regelmäßigen Website-Boosting-Lesern dürfte hinlänglich bekannt sein, dass Duplicate Content aus SEO-Sicht sehr viele Nachteile hat. Insofern ist es wichtig, die Methoden zum Beseitigen von Duplicate Content zu kennen. Eine vergleichsweise neue Möglichkeit stellt nun das Canonical-Tag dar, dessen Wirkungsweise sich am besten an folgendem kleinen Beispiel

nachvollziehen lässt.

Nehmen wir an, der Onlinehändler Herr Müller betreibt einen Shop, in dem man Elektronikartikel kaufen kann. Weil das eingesetzte Shopsystem aus SEO-Sicht einiges falsch macht, werden die URLs der Shopkategorien aufgrund der Klickpfade generiert.

So kann es sein, dass die Shopkategorie, welche alle iPads enthält, auf mehreren Pfaden erreichbar ist. Abbildung 1 veranschaulicht dies.

Je nachdem, über welchen Pfad der Mensch oder eine Suchmaschine also auf die Seite mit den iPads kommt, ist die URL anders. Beide Seiten würden so in die Suchmaschinen-Indizes kommen – ein klassischer Fall von Duplicate Content. Beide Seiten würden so auch externe Links einsammeln. Der interne und externe Linkjuice (positive Wirkung von Links auf Suchmaschinen-Rankings – also quasi die „Linkpower“) müsste durch zwei geteilt werden. Jede Seite würde nur 50 % des möglichen Linkjuices abbekommen.

Der Shopbetreiber Herr Müller möchte nun die URL, welche „apple“ enthält, als die bevorzugte URL benennen, und zwar durch ein Canonical-Tag. Dadurch würde die andere URL aus dem Suchmaschinen-Index entfernt werden und sämtlicher Linkjuice käme der „apple“-Version zugute, wodurch sie viel mehr Potenzial für gute Rankings hätte. Dazu müsste Herr Müller mit einem Canonical-Tag im HTML-Quelltext der anderen Seite (/tablet-pc/ipads) auf diese Seite verweisen. Und zwar so:

```
<link rel="canonical" href="http://www.ein-elektronikshop.de/apple/ipads" />
```

DER AUTOR



Axel Scheuring ist Gründer und Geschäftsführer der eology GmbH.

Zusammen mit seinem 12-köpfigen Team betreut er Kundenwebsites in den Bereichen SEO, PPC, Usability und Social Media.

* siehe Glossar Seite 112-113

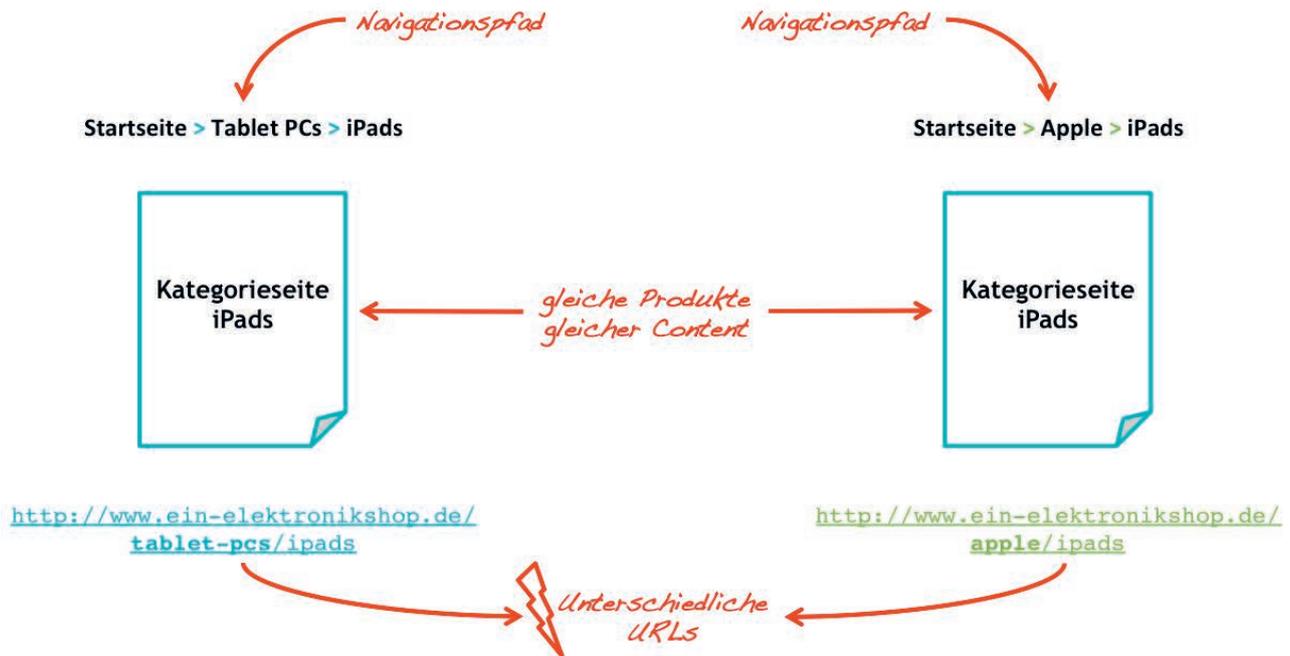


Abb. 1: Duplicate Content entsteht durch unterschiedliche Navigationspfade

Abbildung 2 zeigt, wie das Canonical-Tag in diesem Beispiel gesetzt werden könnte.

Wird das Canonical-Tag korrekt eingesetzt, wird sehr viel Linkjuice weitergegeben. Ähnlich wie bei einem [301 Redirect*](#) gehen nur circa ein bis zehn Prozent des Linkjuices verloren. Im Vergleich zur alten Situation, wo Herrn Müller mit beiden URLs im Index 50 % des

Linkjuices durch die Lappen gingen, also ein echter Gewinn.

Das Canonical-Tag ist also sehr nützlich. Warum ist es dann gefährlich?

Wie Sie gesehen haben, ist das Canonical-Tag sehr einfach einzubauen. Einfach eine Code-Zeile in den <Head>-Bereich der Seite schreiben und schon

läuft's. Korrekterweise müsste man ergänzen: Und schon läuft's richtig oder eben schief. Der größte Vorteil des Canonical-Tags, nämlich das einfache Einbauen, ist gleichzeitig sein größter Nachteil. Viele Website-Betreiber und Programmierer setzen das Canonical-Tag ein, ohne genau zu wissen, wie es eigentlich funktioniert. Dies führt dazu, dass aktuell reihenweise Websites aus

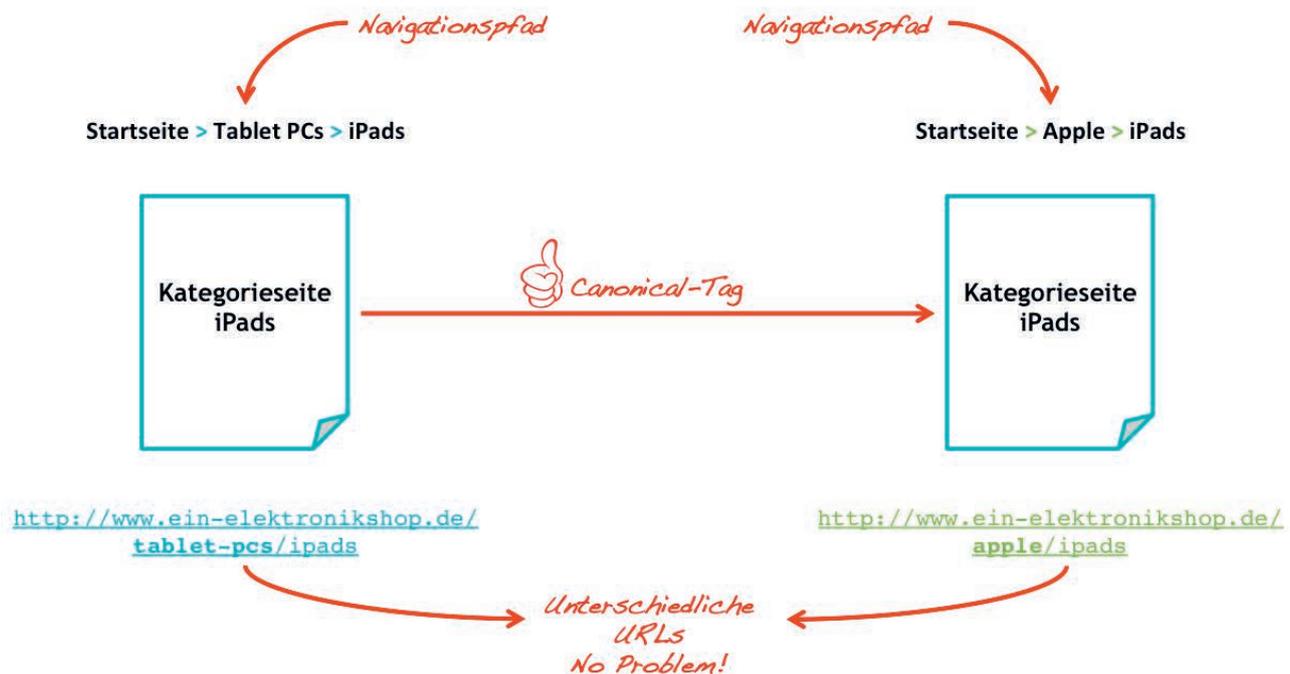


Abb. 2: Durch ein Canonical-Tag wird der Duplicate Content beseitigt

* siehe Glossar Seite 112-113

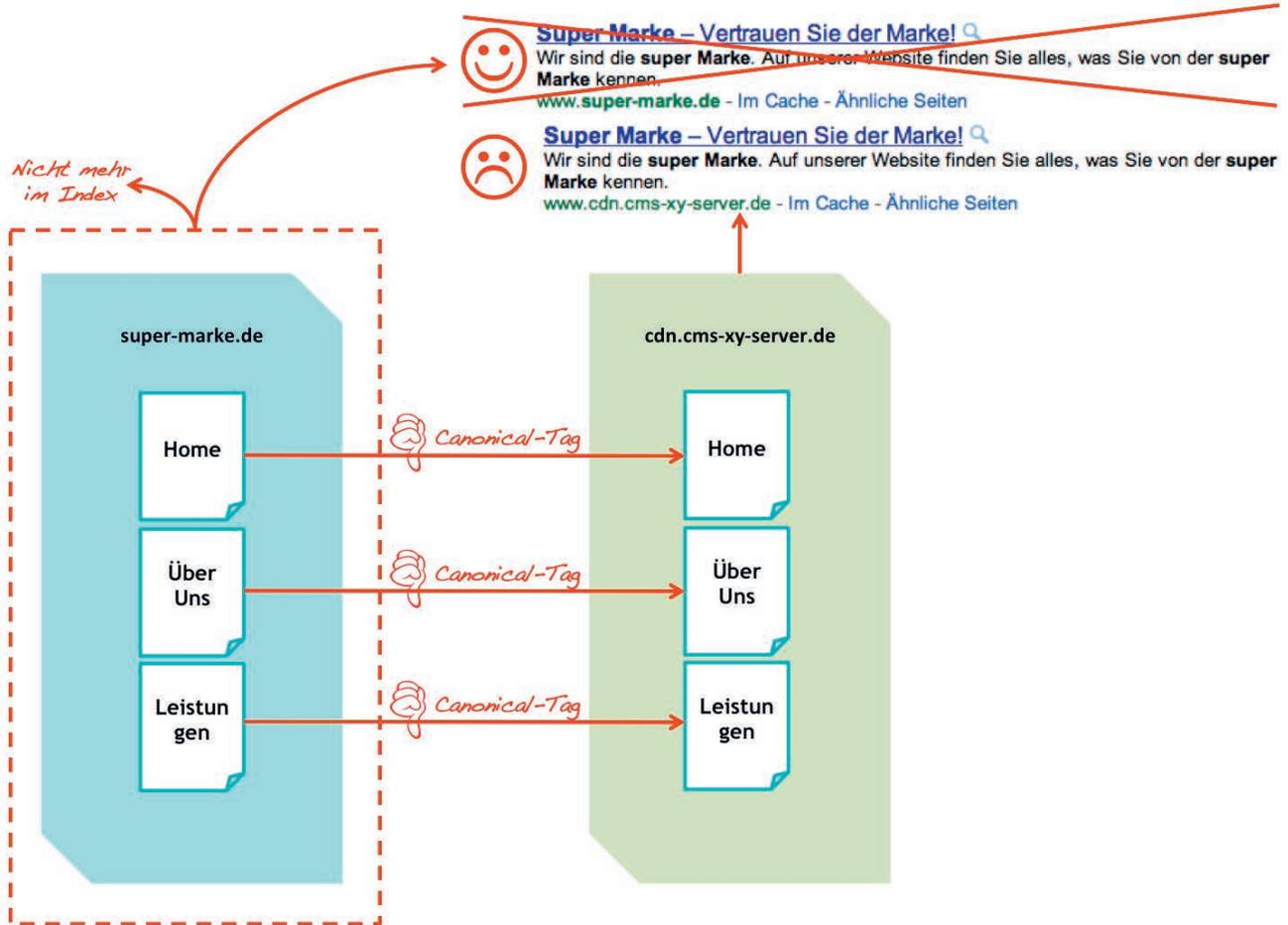


Abb. 3: Durch falsch gesetzte Canonical-Tags kommen reihenweise falsche URLs in den Suchmaschinenindex

Google und anderen Suchmaschinen geschossen werden und im schlimmsten Fall viele bis sämtliche Rankings verlieren.

Weiterhin ist das Canonical-Tag gefährlich, weil man nicht gleich sieht, was damit passiert. Bei Alternativen zum Canonical-Tag, beispielsweise bei serverseitigen 301 Redirects, findet die Weiterleitung für Mensch und

(Such-)Maschine gleichermaßen statt. Der normale Website-Besucher sieht nur noch die Seite, auf die per Redirect weitergeleitet wird, weil ein 301 Redirect auch eine direkte Weiterleitung im Browser verursacht. Wird dagegen ein Canonical-Tag verwendet, sieht der Website-Besucher erst mal noch die alte Version der Seite, denn ein Canonical-Tag leitet nicht im Browser weiter, sondern nur Suchmaschinen-Robots. Ob das Canonical-Tag korrekt gesetzt ist und ob es funktioniert, findet man nur durch ge-

naues Nachsehen im Seiten Quelltext heraus. Sicher sein kann man jedoch erst, wenn dann in den Suchergebnissen bei Google und Co. nicht mehr die alte URL, sondern die neue auftaucht. Insofern werden Fehler beim Einsatz des Tags eventuell erst gar nicht bemerkt. Die böse Canonical-Überraschung kommt dann meist erst, wenn die Site Rankings bei den Suchmaschinen verliert und es schon (fast) zu spät ist.

Sehen wir uns dazu einige typische Canonical-Tag-Fehler an.

Sitewide auf die Startseite

Wenn Sie sehr viele Seiten aus dem Google-Index entfernen möchten, ist das mittlerweile recht einfach. Dafür müssen Sie nur das Canonical-Tag sitewide auf die Startseite setzen. Sitewide heißt, dass von jeder Unterseite ein Canonical-Tag auf die Startseite verweist. Diese Technik wird immer wieder prak-

tiziert, jedoch verständlicherweise meist versehentlich. Häufig wird das Canonical-Tag auf den einzelnen Seiten automatisiert durch ein Content-Management-System generiert. Dabei kommt es leider oft vor, dass hier irgendetwas schief läuft und das Canonical-Tag nicht auf die korrekte Originalversion der Seite verweist, sondern auf die Startseite. Das kann beispielsweise passieren, wenn das URL-Matching aus technischen Gründen nicht klappt und somit der URL-Teil nach dem Domain-Namen einfach abgeschnitten wird. Das Canonical-Tag verweist also anstatt auf die Deep-URL <http://eine-website.de/unterseite.html> auf die Startseite der Domain <http://eine-website.de/>.

Nachdem Google aktuell nach eigenen Aussagen zwei von drei Empfehlungen durch ein Canonical-Tag tatsächlich befolgt, ist die Wahrscheinlichkeit recht groß, dass fast alle Seiten – außer der

Startseite – aus dem Google-Index fliegen. Man wird also deutlich schlechter gefunden.

Canonical-Tags auf falsche URLs

Ein weiterer unangenehmer Canonical-Fehler ist, dass das Tag auf falsche URLs verweist. So könnte es z. B. sein, dass es versehentlich auf eine nicht existente URL (404-Fehlerseite) zeigt. Das wäre jedoch noch der harmloseste Fall. Die Suchmaschinen würden es in diesem Fall einfach ignorieren und das Duplicate-Content-Problem bestünde weiterhin.

Der wesentlich ärgerlichere Fall wäre, dass das Canonical-Tag auf eine falsche URL verweist, die tatsächlich existiert. Es gibt beispielsweise Fälle, in denen eine falsche Domain sitewide im Canonical-Tag steht. Sehen wir uns dazu das Beispiel aus Abbildung 3 an.

Die Domain „super-marke.de“ ranked vor dem Einsatz des Canonical-Tags ganz normal für viele interessante Suchbegriffe.

Wird nun durch einen Fehler im Content-Management-System beispielsweise die Subdomain „cdn.cms-xy-server.de“ aus einem Content Delivery Network ins Canonical-Tag generiert, löst diese dann logischerweise relativ schnell die „super-marke.de“ im Ranking ab. Das Canonical-Tag funktioniert nämlich auch domainübergreifend hervorragend. Das ist aber in diesem Fall gerade aus Markenbildungssicht schlecht, denn man präsentiert sich so in den SERPs nicht mehr als die bekannte Marke mit super Markendomain, sondern als irgendetwas, das kein Mensch kennt und wahrscheinlich auch nicht anklicken will. Die Suchenden werden sich fragen: Ist das wirklich die Supermarke, die ich gesucht habe? Wenn ich mir so die Adresse „cdn.xy-server.de“ ansehe, glaube ich das nicht. Lieber nicht darauf klicken. Wer weiß, was sich dahinter verbirgt ...

Canonical-Tag vergessen

Gerade beim Relaunch einer Website, bei dem sich die URLs der einzelnen Seiten ändern, ist es wichtig, diese suchmaschinenfreundlich auf die neuen Seiten weiterzuleiten. Das muss sein, weil die alten Seiten jahrelang interne und externe Links eingesammelt haben, die sonst ins Leere laufen würden, und weil sie sich in den Augen der Suchmaschinen ein gewisses Vertrauen erarbeitet haben. Die alten Seiten gibt es ja in dieser Form dann gar nicht mehr. Die Rankings würden verloren gehen. Dafür bieten sich 301 Redirects an. Falls das serverseitig nicht möglich ist, kann man die Rankings auch mit dem Canonical-Tag retten.

Leider kommt es zu oft vor, dass in der Hektik eines Relaunches schlicht vergessen wird, die Canonical-Tags von den alten auf die neuen URLs zu setzen, obwohl man es ursprünglich vielleicht sogar auf der Agenda hatte. Es sei aber noch einmal eindringlich darauf hingewiesen, dass das saubere Weiterleiten von URLs bei einem Website-Relaunch kein überflüssiges SEO-Gespinnne ist, sondern im Gegenteil essenziell, wenn auch nach dem Relaunch noch Traffic über Google und andere Suchmaschinen kommen soll.

Da geht noch mehr – weitere Probleme mit dem Canonical-Tag

Das Canonical-Tag steckt voller Überraschungen, die noch weit über die bereits erläuterten „Klassiker“ hinausgehen. Eine kleine Portion Ungeschick reicht bereits aus, um noch viel mehr ungeplante Ergebnisse damit zu produzieren.

Canonical-Tag vice versa

Was passiert wohl, wenn das Canonical-Tag von Seite 1 auf Seite 2 verweist und das Canonical-Tag von Seite 2 wieder auf Seite 1?

Dieser Frage ging auch kürzlich Dr.

Peter Meyers nach, als er in einem Experiment sechs extreme Canonical-Tricks testete (www.seomoz.org/blog/6-extreme-canonical-tricks).

In jedem Fall ist eine Menge unkontrollierter Canonical-Tag-Spaß zu erwarten. Im Experiment tauchten beide Seiten im Google-Index auf. Das Duplicate-Content-Problem war also nicht gelöst. Mehr noch: Der Titel der 2. Seite wurde auf der Google-Suchergebnisseite nicht im Original ausgespielt, sondern auf den jeweiligen Suchbegriff angepasst. So hat man nicht mehr die Kontrolle über seine Website-Präsentation in den Suchergebnissen, was ja eigentlich sehr wichtig ist.

Doppelt gesetzte Canonical-Tags

Gelegentlich wird das Canonical-Tag auf einer Seite auch versehentlich doppelt gesetzt. Meist passiert das, wenn ein Content-Management-System im Spiel ist, welches Canonical-Tags automatisiert generiert. Meist hat das auch negative Auswirkungen für die Website.

Beispielsweise kann das passieren, wenn man das Content-Management-System „WordPress“ einsetzt. Läuft das Plugin „All in one SEO“ mit der Standardeinstellung „Canonical URLs“, erzeugt dieses automatisch auf jeder Seite ein Canonical-Tag, welches jeweils auf die aktuell angezeigte Seite verweist.

Setzt man zusätzlich noch das Plugin „Canonical URL's“ von Joost de Valk ein, kann es zu Doppelgenerierungen von Canonical-Tags kommen. Mit dem letztgenannten Plugin kann man nämlich Canonical-Tags manuell setzen. Dies bietet sich an, wenn man beispielsweise einen Artikel in zwei unterschiedlichen Blogs veröffentlichen und dabei Duplicate Content vermeiden möchte.

Sehen wir uns ein konkretes Beispiel an, wo es zu Doppelsetzungen von Canonical-Tags kommen kann.

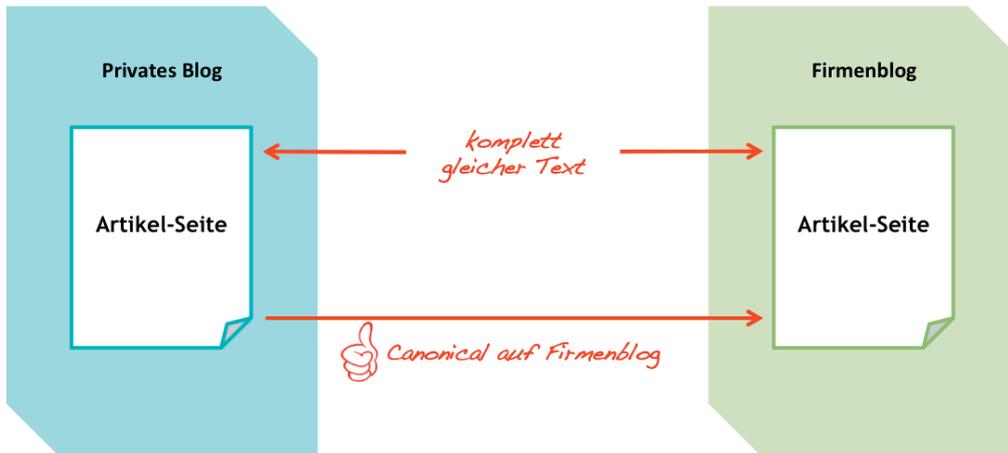


Abb. 4: Ein domainübergreifendes Canonical-Tag entfernt Duplicate Content

Sagen wir, Sie bloggen für Ihre Firma im Firmen-Blog und betreiben zusätzlich noch ein privates Blog. Nun schreiben Sie einen Artikel, den Sie im Firmenblog veröffentlichen, und finden, er passt auch ganz gut in Ihr privates Blog. Würden Sie den Artikel auf beiden Seiten veröffentlichen, hätten Sie ein Problem: Duplicate Content, und zwar domainübergreifend. Schließlich würde sich ein und derselbe Artikel auf zwei unterschiedlichen Domains befinden. Er wäre also unter zwei unterschiedlichen URLs erreichbar.

Vor einiger Zeit hätten Sie deshalb den Artikel vielleicht nicht zusätzlich

in Ihrem privaten Blog veröffentlicht. Da es aber jetzt das Canonical-Tag gibt und es auch domainübergreifend funktioniert, können Sie den Artikel getrost auf beiden Sites veröffentlichen, solange Sie mit dem Canonical-Tag von einer auf die andere verweisen. Wenn Sie z. B. möchten, dass nur der Artikel auf dem Firmenblog bei Google und anderen Suchmaschinen auffindbar ist, setzen Sie das Canonical-Tag von der Artikel-Seite im privaten Blog auf die entsprechende Seite im Firmenblog, so wie in Abbildung 4 beschrieben.

Damit wäre für Suchmaschinen klar, welche Variante gelistet werden

soll. In diesem Fall würde die Artikel-Seite auf dem Firmenblog übrigens auch von Links profitieren, die jemand auf die Artikel-Seite Ihres privaten Blogs setzt.

Im Beispiel erzeugt das Content-Management-System nun aber zwei Canonical-Tags, was Abbildung 5 veranschaulicht. Vielleicht wurde auch im Beispiel eine ungünstige Kombination aus mehreren Plugins eingesetzt. Jedenfalls verweist eines auf diese Seite selbst, während das andere eigentlich korrekt auf das Firmenblog zeigt, aber weiter unten im Quelltext steht. Dadurch wird es von den Suchmaschinen

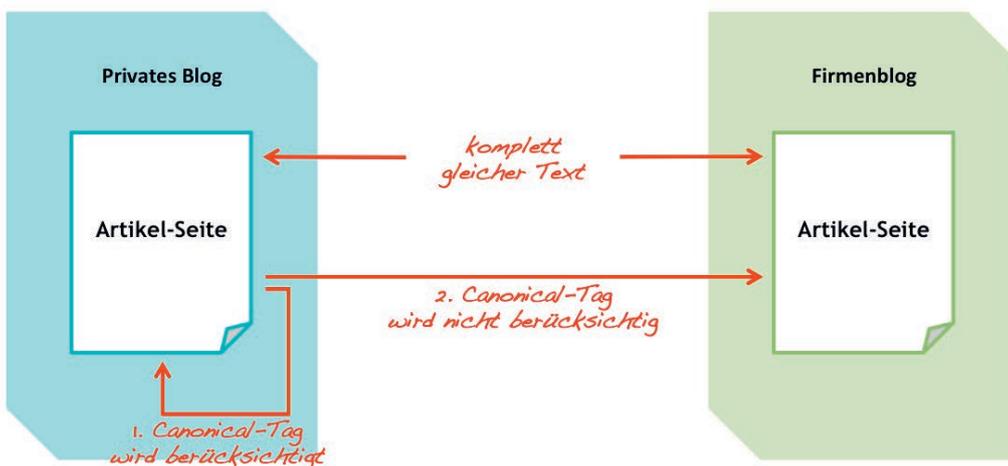


Abb. 5: Bei zwei Canonical-Tags wird nur das berücksichtigt, welches im Quelltext weiter oben steht

nicht berücksichtigt.

Hier im Beispiel (Abb. 5) bleibt der Duplicate Content also weiter bestehen, da beide Seiten nach wie vor im Index sind.

Im falschen HTML-Bereich

Gelegentlich sieht man auch Websites, die das Canonical-Tag versehentlich nicht im <Head> der Seite einbauen, sondern in einem anderen HTML-Bereich. Selbst wenn das Tag syntaktisch völlig korrekt ist und die angegebene URL tatsächlich existiert, wird es so nicht funktionieren. Das bestätigt auch Matt Cutts von Google: „First off, here's a thought exercise: should Google trust rel=canonical if we see it in the body of the HTML? The answer is no, because some websites let people edit content or HTML on pages of the site. If Google trusted rel=canonical in the HTML body, we'd see far more attacks where people would drop a rel=canonical on part of a web page to try to hijack it“ (Quelle: www.matcutts.com/blog/rel-canonical-html-head/).

Google berücksichtigt also keine Canonical-Tags, die im <Body> der Seite stehen, und zwar aus Sicherheitsgründen: Bei Websites, die User Generated Content erlauben, könnte es sonst passieren, dass Inhalte dieser Websites von einigen Nutzern auf andere Seiten weitergeleitet werden.

Auch ein Canonical-Tag in einem Pseudo-<Head> hätte damit keine Wirkung, d. h. das Konstrukt aus Listing 1 wäre nicht korrekt. So könnte man also keinen Duplicate Content beseitigen.

```
<html>
<head>
  <title>Canonical-Tag funktioniert
    nicht</title>
</head>
<body>
  <p>Das Canonical-Tag kann so nicht funk-
    tionieren, weil es nicht im korrekten
    Head-Bereich steht</p>
  <head>
    <link rel="canonical"
      href="http://www.eine-website.
      de/korrekte-url.html" />
  </head>
</html>
```

Deshalb sollte man genau darauf achten, das Canonical-Tag an der richtigen Stelle einzubauen.

Yahoo- und Bing-Problematik

Bei Google werden Empfehlungen durch das Canonical-Tag sehr stark berücksichtigt. Weil Google in Deutschland einen Marktanteil von circa 90 Prozent hat, kann man mit Hilfe des Canonical-Tags also sehr viele Duplicate-Content-Probleme in den Griff bekommen – aber nicht alle, denn es gibt es auch noch andere Suchmaschinen wie z. B. Bing und Yahoo, bei denen das Canonical-Tag leider sehr viel schlechter funktioniert. Empfehlungen durch ein Canonical-Tag werden bei diesen beiden Suchmaschinen aktuell nur manchmal berücksichtigt, in jedem Fall aber sehr viel seltener als bei Google. Wenn man bedenkt, dass Bing und Yahoo zusammen immerhin auf einen Marktanteil von fünf Prozent kommen, würde das im Extremfall bedeuten, dass man fünf Prozent seiner Duplicate-Content-Probleme nie lösen könnte, wenn man auf das Canonical-Tag setzt. Das ist sehr viel. Insofern sollte man immer überlegen, welche Alternativen möglich sind, z. B. 301 Redirects.

Canonical-Tag: Make or Break?

Ist das Canonical-Tag nun eine großartige Neuerung für alle Duplicate-Content-geplagten Webmaster oder ist es viel zu gefährlich und nutzlos, weil es doch auch Alternativen gibt?

Wie so oft im SEO-Bereich kann man auch hier nur antworten: „It depends!“ Generell sollte Duplicate Content durch eine gut durchdachte Site-Architektur gar nicht erst entstehen. Wenn es aber doch passiert, sollte man dem Canonical-Tag immer die Alternativen 301 Redirect oder beispielsweise „NoIndex, Follow“ auf den Duplikaten vorziehen. Bei parameterbedingtem Duplicate Content kann man auch in den Webmaster-Tools die Parameter für Suchmaschinen sperren. Falls aus technischen Gründen eine serverseitige Lösung mit 301 Redirects nicht möglich ist, darf bzw. sollte das Canonical-Tag genutzt werden. Aber eben auch dann, wenn beispielsweise unterschiedliche Navigationspfade zu unterschiedlichen URLs führen. Denn hier möchte man normalerweise ja nicht gleich einen Redirect-bedingten URL-Refresh erzeugen.

Es gibt also durchaus sinnvolle Einsatzmöglichkeiten für das Canonical-Tag, welches mit Sicherheit eine der bedeutendsten Erleichterungen im SEO-Bereich der letzten Jahre ist. Schließlich leistet es bereits vielen Website-Betreibern schon seit längerem wertvolle Dienste im ernstzunehmenden Kampf gegen Duplicate Content Probleme.¶