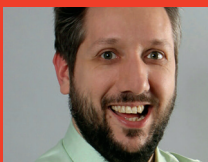




Fili Wiese

ALLES, WAS SIE FÜR DIE UMSTELLUNG AUF HTTPS WISSEN MÜSSEN (TEIL 1)

DER AUTOR



Fili Wiese ist ein renommierter SEO-Spezialist und hat früher in leitender Funktion im Google Search Quality Team mitgearbeitet. Bei SearchBrothers.com geht er mit Erfolg gegen die Abstrafung von Websites durch Google-Penalties vor und bietet SEO-Consulting mit SEO-Audits und SEO-Workshops.

Unter dem Eindruck der Enthüllungen von Edward Snowden und angesichts der Google-Initiative, mehr Websites zur Umstellung zu bewegen, wechseln immer mehr Websites zu HTTPS. Die Umstellung auf HTTPS ist eine technische Herausforderung, die man nicht unterschätzen sollte. Unter dem Aspekt der Suchmaschinenoptimierung muss man Ressourcen bereitstellen, langfristig planen, gründlich vorbereiten und den Umzug konsequent durchziehen; Risiken sind dabei generell nicht auszuschließen. Dieser Leitfaden erklärt Anfängern wie auch erfahrenen Website-Betreibern, was Sie hinsichtlich der Suchmaschinenoptimierung beim Wechsel von HTTP zu HTTPS beachten müssen, warum der Wechsel zu HTTPS so

wichtig ist, welches SSL-Zertifikat für Sie das richtige ist und ob Sie Ihre Website in einem Durchgang umstellen sollten; außerdem erfahren Sie mehr über die häufigsten technischen On-Page-Signale, die auf HTTPS aktualisiert werden müssen, damit keine widersprüchlichen Signale an die Suchmaschinen geschickt werden. Der SEO-Experte und ehemalige Googler aus dem Search Quality Team Fili Wiese erklärt Ihnen umfassend, wie man die Google Search Console für die Umstellung auf HTTPS konfiguriert, die Wirkung der Umstellung mithilfe der Google Search Console und der Server-Logfiles überprüft und die gesamten Performance-Signale zum Anwender mit HSTS und HTTPS/2 verbessert.

Foto: Imilian / thinkstockphotos.de

Was ist HTTPS und warum sollte ich mir Gedanken darüber machen?

Wenn Sie eine Website besitzen oder Websites im Internet besuchen, geht kein Weg daran vorbei: Sie müssen sich Gedanken über HTTPS machen. HTTPS steht für „Hypertext Transfer Protocol over TLS“ und ist ein Protokoll für die sichere Kommunikation zwischen Computernetzwerken wie beispielsweise dem Browser auf einem lokalen Computer oder dem Server, der die Inhalte zur Verfügung stellt, auf die Sie zugreifen. Jede Website im „World Wide Web“ verwendet entweder HTTP oder HTTPS. Die folgende Website beispielsweise verwendet das HTTPS-Protokoll (sichtbar in der Adressleiste).

HTTPS hat zahlreiche Vorteile, aber für die meisten Benutzer zählen vor allem folgende Aspekte:

- » Datensicherheit: HTTPS verhindert beispielsweise Janusangriffe.
- » Datenschutz: HTTPS verhindert beispielsweise Online-Lauschangriffe durch Dritte.
- » Geschwindigkeit (dazu später mehr).

Auch die Betreiber einer Website profitieren in vieler Hinsicht von HTTPS:

- » Datensicherheit für die Verarbeitung kritischer Informationen, beispielsweise beim Zahlungsverkehr.
- » Die Referrer-Informationen bleiben in Analytics erhalten, im Unterschied zu HTTP-Seiten: Kommt ein Besucher von einer HTTPS-Website aus zu einer HTTP-Website, so wird der Referrer nicht übermittelt. Eine HTTPS-Website hingegen behält die Referrer-Informationen, gleich ob der Besucher über eine HTTP- oder HTTPS-Website zur aktuellen Seite gelangt ist.
- » Potenziell verbessertes Ranking in den Google-Suchergebnissen.

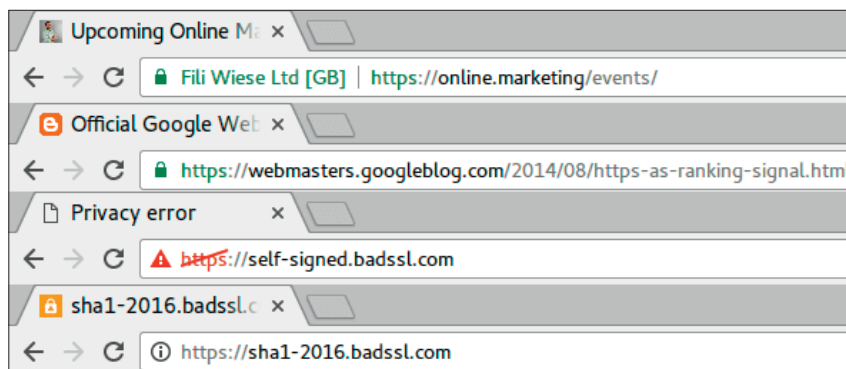


Abb. 1: Google Chrome zeigt verschiedene HTTPS-URLs an

- » HTTP-Websites erscheinen bei zukünftigen Browser-Updates als unsicher, HTTPS-Seiten hingegen als sicher.
- » Geschwindigkeit (dazu später mehr).

HTTPS ist jedoch nicht ganz einfach zu handhaben und deshalb verwendet bisher nur rund die Hälfte der Websites im Internet dieses sichere Protokoll, wobei die Anzahl ständig zunimmt. Unter anderem sind folgende Probleme zu bewältigen:

- » Zusätzliche Kosten, weil kommerzielle SSL-Zertifikate kostenpflichtig sind; außerdem ist eine moderne Server-Infrastruktur erforderlich, damit keine Kosten für RAM/CPU entstehen.
- » Technische Herausforderungen, denn die Implementierung von SSL-Zertifikaten auf einem Server ist alles andere als einfach; bis vor Kurzem musste man für eine HTTPS-fähige Domain eine eindeutige IP-Adresse haben, aber glücklicherweise gibt es jetzt die Server Name Indication, die von fast allen wichtigen Browsern unterstützt wird.
- » Die Suchmaschinen betrachten die Umstellung von HTTP auf HTTPS als Verschiebung von Inhalten, was der Website bis zum erneuten Crawl und Verarbeiten aller neuen HTTPS- und alten HTTP-URLs ein niedrigeres Ranking einbringt.
- » Konflikte mit dem ursprünglichen Design des World Wide Web, weil das zusätzliche „S“ im HTTPS-Pro-

tokoll die korrekte Funktion der Hyperlinks unterbindet – und die sind ein fundamentaler Stützpfiler des World Wide Web.

- » Die Umstellung auf HTTPS ist eine Langzeitstrategie. Sobald man sich auf sie eingelassen hat, kommt man praktisch nicht daran vorbei, eine HTTPS-Variante der Website zu betreiben. Selbst wenn die HTTP-Variante den Datenverkehr nur zur HTTP-Variante zurückleitet, muss sie dennoch aktiv gehalten werden, damit sie weiterhin externen Linkjuice und Besucher zur HTTP-Variante weiterleitet. Anders ausgedrückt: Sobald eine Website live ist und unter HTTPS betrieben wird – sei es auch nur für kurze Zeit –, wird man voraussichtlich während der Lebensdauer der Website weder die HTTPS- noch die HTTP-Variante abschalten können.

Dennoch überwiegen die langfristigen Vorteile. Hauptthema dieses Leitfadens ist das Problem der Migration von Inhalten unter dem Blickwinkel der Suchmaschinenoptimierung.

Voraussetzung: SSL-Zertifikate

Bevor man sich näher mit den Aspekten der Suchmaschinenoptimierung bei der Umstellung auf HTTPS befassen kann, muss man prüfen, ob der Server richtig eingestellt ist und der Migration der Inhalte nichts im Weg steht.

Kaufen Sie ein öffentliches SSL-Zertifikat

Obwohl man selbst erstellte SSL-Zertifikate oder kostenlose von der Community zur Verfügung gestellte SSL-Zertifikate verwenden kann und auch wenn hinter den meisten Arten öffentlicher SSL-Zertifikate Zertifizierungsstellen stehen, bieten ausschließlich öffentliche Zertifikate einen Zusatznutzen, nämlich SSL-Zertifikate mit erweiterter Validierung (EV). Für diese EV-SSL-Zertifikate wird die Identität des Antragstellers gründlicher überprüft, und es kann aufwendiger sein, die Genehmigung für die Ausgabe zu erwirken. Der Besucher der Website erkennt das öffentliche SSL-Zertifikat an dem grünen Balken mit dem Firmennamen in der Adresszeile des Browsers.

Bei der Auswahl eines bestimmten SSL-Zertifikats gilt es zu bedenken, dass von der Community zur Verfügung gestellte SSL-Zertifikate noch in den Kinderschuhen stecken und die Preise für kommerzielle SSL-Zertifikate während der letzten Jahre auf unter zehn US-Dollar pro Jahr und SSL-Zertifikat gefallen sind. Daher ist für kommerzielle Websites vorerst die Nutzung öffentlicher SSL-Zertifikate zu empfehlen; später kann man immer noch auf andere Optionen umschwenken.

Bei der Auswahl eines öffentlichen SSL-Zertifikats ist auch zu beachten, dass man verschiedene Validierungsarten von SSL-Zertifikaten kaufen bzw. verwenden kann. Nach meiner Erfahrung bei Google ist es eigentlich egal, für welches Zertifikat Sie sich entscheiden; jedes SSL-Zertifikat ist geeignet, aber für den Benutzer der Website kann es einen Unterschied machen.

Verschlüsselung

Bei der Erstellung eines (kommerziellen oder selbst erstellten) SSL-Zertifikats gibt es verschiedene Optionen. Es ist besser, ein SHA2-Zertifikat zu wählen (z. B. SHA256), da es sicherer

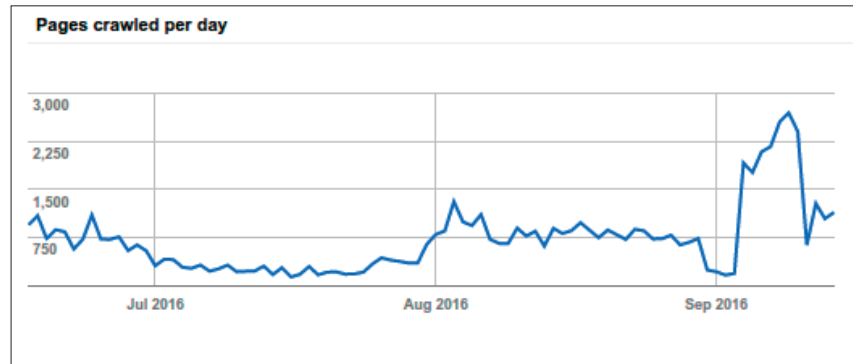


Abb. 1: Google Chrome zeigt verschiedene HTTPS-URLs an

ist als ein SHA1-Zertifikat. SHA1-Zertifikate wurden aus diesem Grund von den meisten der verbreiteten Browser abgelehnt. Ende des Jahres 2016 werden Websites, die SHA1-Zertifikate verwenden, als unsicher betrachtet werden; das würde dem Zweck der Nutzung von HTTPS entgegenstehen.

Implementierung auf dem Server

Um die SSL-Zertifikate auf der Server-Infrastruktur zu implementieren, kontaktieren Sie den Hosting-Provider, das IT-Team oder die Web-Developer.

Wenn Sie Server Name Indication verwenden, prüfen Sie in der Datenanalytik der aktuellen Website, ob sie noch häufig von bestimmten alten Browsern aus besucht wird, die SNI nicht unterstützen.

Nach der Einrichtung des Servers mit einem SSL-Zertifikat für den Domain-Namen auf Port 443 müssen die Einrichtung und die Server-Umgebung geprüft und validiert werden. Zur Validierung des SSL-Zertifikats kann das SSL Shopper Tool verwendet werden. Zur Prüfung der Server-Einstellung können Sie das SSL Labs Tool einsetzen. Eventuelle Fehler müssen behoben werden, bevor Sie fortfahren.

Hinweis: Zur Vermeidung von Problemen beim Update der Website für die in den nächsten Schritten durchgeführte Umstellung auf HTTPS ist es empfehlenswert, ein eigenes Hauptverzeichnis auf demselben Server oder einer anderen Server-Instanz einzurichten und den HTTPS-Verkehr dorthin weiterzuleiten.

Vorbereitung der Umstellung auf HTTPS

Dieser Leitfaden geht davon aus, dass folgende Voraussetzungen erfüllt sind, bevor die nächsten Schritte erfolgen können:

- » Am Inhalt wurden außer den Link-Annotationen keine Veränderungen vorgenommen.
- » An den Dokumentenvorlagen wurden außer den Link-Annotationen keine Veränderungen vorgenommen.
- » An der URL-Struktur wurden außer der Protokolländerung keine Veränderungen vorgenommen.
- » Die HTTPS-Variante des Domain-Namens ist live auf Port 443 und wurde wie vorstehend beschrieben validiert.

Das Hauptverzeichnis des Domain-Namens auf HTTPS wird höchstwahrscheinlich ein leeres Verzeichnis-Listing anzeigen.

Wenn eine der drei erstgenannten Annahmen nicht erfüllt ist, sollten Sie auf jeden Fall den Beitrag „Wie Sie den Inhalt Ihrer Website verschieben“ auf dem offiziellen Google Webmaster-Blog lesen.

Definition einer Strategie für das Verschieben der Inhalte

Als nächsten Schritt gilt es, eine Strategie für die Umstellung auf HTTPS zu wählen. Je nachdem, ob man eine kleine Website (z. B. weniger als zehntausend URLs) oder eine große Website (z. B. über eine Million URLs)

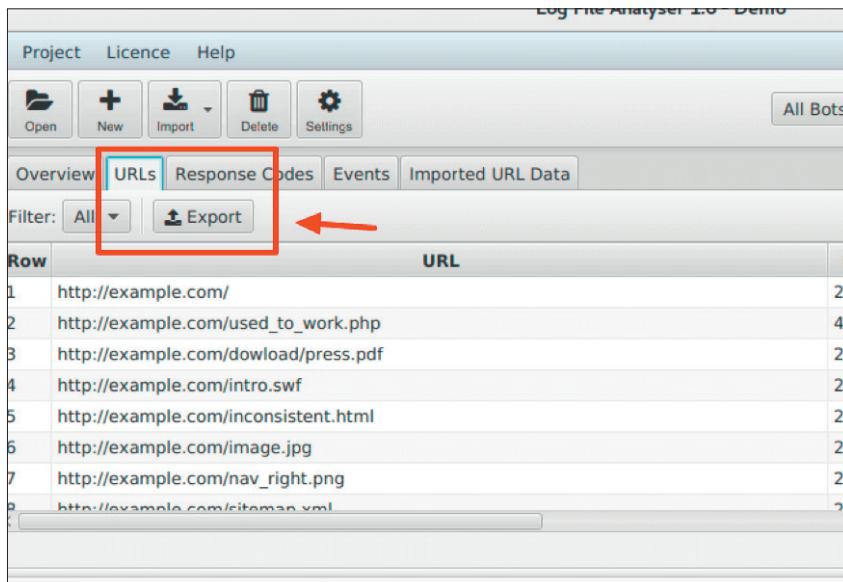


Abb. 3: Beispiel für den Export von URLs von Server-Logs mithilfe des Screaming Frog Log Analyser

auf HTTPS umstellt, ergeben sich unterschiedliche Möglichkeiten für das Verschieben der Inhalte, nämlich:

- » Man stellt die gesamte Domain einschließlich aller Subdomains auf einmal auf HTTPS um.
- » Man stellt erst nur eine oder mehrere Subdomains und/oder Unterverzeichnisse auf HTTPS um und holt den Rest dann nach.
- » Man verschiebt erst nur den Inhalt und betreibt zwei Duplikate unter HTTP und HTTPS, bevor man die endgültige Umstellung vornimmt.

Im Rahmen dieser Strategie gilt es auch, die folgende Frage zu beantworten: *Wie lange kann man noch auf die HTTP-Variante zugreifen?*

Je nach Größe der Website, Verfügbarkeit des IT-Support-Teams und der Organisationsstruktur hinter der Website (z. B. interne Firmenrichtlinien) sind hier unterschiedliche Faktoren zu berücksichtigen. Der vorliegende Leitfaden kann zwar zu den letzten beiden Punkten nicht für jede Website eine Aussage treffen, aber der erste Punkt ist definitiv ein Faktor, der für die Suchmaschinenoptimierung relevant ist – und damit für das verfügbare Crawl-Budget.

Der Stellenwert des Crawl-Budgets

Damit Suchmaschinen die Protokollierung verarbeiten können, müssen ihre Bots einen Großteil der HTTP-URLs sowie alle neuen HTTPS-URLs der Website erneut crawlen. Für eine Website mit 1.000.000 URLs müssen die Suchmaschinen-Bots also mindestens um die 2.000.000 URLs (oder einen großen Teil davon) erneut crawlen, um die 301-Weiterleitungen zu finden und die Rankings für die neuen HTTPS-URLs auf der Grundlage der historischen HTTP-URLs erneut zu berechnen. Ein Suchmaschinen-Bot, der pro Tag im Durchschnitt 30.000 eindeutige No-repeat-URLs auf der Website crawlt, braucht rund 67 Tage für das erneute Crawlen aller URLs. Während dieser Zeit kann die Website im Suchmaschinen-Ranking abrutschen, sofern es keine „im Crawl-Budget vergeudeten URLs“ gibt.

Nutzen Sie die Server-Logfiles

Um sicherzustellen, dass Suchmaschinen-Bots kein Crawl-Budget vergeuden, prüfen Sie zunächst gründlich die Server-Logfiles und stellen fest, welche URLs während der letzten beiden Jahre (falls möglich auch länger) von jedem Suchmaschinen-Bot gecrawlt wurden. Es ist auch hilfreich zu wissen, wie oft jede

URL gecrawlt wurde, um die Priorität festzustellen, aber dafür braucht man sowieso alle URLs. Dieser Leitfaden konzentriert sich vor allem auf Crawl-Daten von Googlebot.

Für kleinere Websites lässt sich diese Aufgabe recht einfach mit dem Screaming Frog Log Analyser erledigen. Für eine größere Website sprechen Sie das IT-Team an und/oder verwenden Sie eine Big-Data-Lösung wie beispielsweise Google BigQuery, um alle URLs zu extrahieren.

Möglicherweise müssen Sie auch die Logdateien beim Hosting-Anbieter der Website anfordern. Wenn keine Logdateien für die letzten beiden Jahre vorliegen (und die Website nicht brandneu ist), beginnen Sie so schnell wie möglich mit dem Loggen. Ohne Logdateien fehlen dem Unternehmen wichtige und entscheidende Daten zur Suchmaschinenoptimierung sowie ausschlaggebende analytische Geschäftsdaten.

Speichern Sie die extrahierten URLs in einer eigenen Datei (eine URL pro Zeile), z. B. als logs_extracted_urls.csv.

Extrahieren Sie Sitemap-URLs

Unter der Voraussetzung, dass eine Website eine oder mehrere XML-Sitemaps hat und diese Sitemaps alle eindeutigen kanonischen URLs der indexierbaren Seiten der Website beinhalten, zeichnet die Google Search Console auf, wie viele URLs der XML-Sitemaps Google aktuell vorgelegt werden. Laden Sie alle eindeutigen URLs der XML-Sitemaps herunter und extrahieren Sie sie.

Speichern Sie die extrahierten URLs in einer eigenen Datei (eine URL pro Zeile), z. B. als logs_extracted_urls.csv.

Welcher Anteil der Website soll umgestellt werden?

In dieser Phase liegen genügend Daten vor, um über den nächsten Schritt zu entscheiden: Welcher Anteil der Web-

site soll auf HTTPS umgestellt werden?

Um die Zahl der URLs für die Umstellung zu berechnen, müssen folgende Daten vorliegen:

- » Eine Liste der von Googlebot gecrawlten eindeutigen URLs, extrahiert aus den Server-Logfiles.
- » Die durchschnittliche Anzahl der von Googlebot pro Tag gecrawlten URLs, beruhend auf Zahlen der Crawl-Statistiken in der Google Search Console und den Logdateien vom Server.
- » Eine Liste der aus den XML-Sitemaps extrahierten eindeutigen URLs.

Geht aus den Daten aus den Server-Logfiles hervor, dass Googlebot pro Tag im Durchschnitt zwischen 5 % und 100 % der URLs crawlt, verglichen mit der Gesamtzahl der aus den XML-Sitemaps extrahierten eindeutigen URLs, so kann man es durchaus wagen, die gesamte Domain auf einmal auf HTTPS umzustellen. Bei diesem Szenario hat man gute Chancen, dass die Suchmaschinen-Bots die gesamte Domain innerhalb der nächsten drei bis vier Wochen erneut crawlen – abhängig von der internen Link-Struktur und einigen anderen Faktoren. Bezeichnen wir diese Situation als Szenario 1, **„Umstellung in einem Durchgang“**.

Liegt die durchschnittliche Anzahl der von Googlebot gecrawlten URLs jedoch zwischen 1 % und 5 % verglichen mit der Gesamtzahl der aus den XML-Sitemaps extrahierten eindeutigen URLs, sollte man auf Nummer sicher gehen und etappenweise je eine oder mehrere Subdomains und/oder Unterverzeichnisse auf HTTPS umstellen. Bei diesem Szenario kann es nämlich möglicherweise lange dauern, bis Googlebot alle URLs der gesamten Domain erneut gecrawlt hat, sodass mehr als die üblichen paar Wochen vergehen, bis Ihre Website ihr früheres Ranking in den Google-Suchergebnissen wiedererlangt. Dies ist Situation 2, **„Umstellung in Etappen“**. Hierbei werden alle Subdomains und

Unterverzeichnisse nach und nach umgestellt, bis die gesamte Domain nach HTTPS umgezogen ist.

Für richtig große Website gibt es eine dritte Option, nämlich den parallelen Betrieb von zwei Websites, eine unter HTTP, die andere unter HTTPS. Während man auf das Recrawlen einer ausreichend großen Zahl von URLs wartet, wird der Inhalt mithilfe kanonischer URLs verschoben. Damit das funktioniert, müssen sich die Website-Betreiber darauf verlassen, dass die Suchmaschinen den kanonischen Signalen der Website vertrauen. Dies ist Szenario 3, **„Umstellung mithilfe kanonischer URLs“**. Sobald genug URLs erneut gecrawlt wurden, kann man den Umzug nach HTTPS nach Szenario 1 oder 2 abschließen.

Crawlen der HTTP-Website

Setzen Sie als Nächstes einen Crawler wie z. B. Screaming Frog SEO Spider, DeepCrawl, Botify oder OnPage.org ein, um die gesamte Website oder die wichtigsten Teile der aktuellen Website im HTTP-Protokoll zu crawlen. Extrahieren Sie alle eindeutigen URLs, die von den Suchmaschinen gecrawlt werden können; sie sind intern miteinander verbunden. Dazu gehören alle internen Inhalte der Website wie robots.txt, JavaScript, Image, Font und CSS-Dateien. Sie benötigen diese Daten, damit Sie am Ende erneut prüfen können, ob die Umstellung erfolgreich war.

Speichern Sie die extrahierten URLs in einer eigenen Datei (eine URL pro Zeile), z. B. als `logs_extracted_urls.csv`.

Hinweis: Für sehr große Websites von z. B. mehr als zehn Millionen URLs ist entweder das Szenario „Umstellung in Etappen“ oder „Umstellung mithilfe kanonischer URLs“ als sicherste Vorgehensweise zu empfehlen. Versuchen Sie, die Website in handliche Abschnitte aufzuteilen, entweder nach Subdomains und/oder nach Unterverzeichnissen, und crawlen Sie die Abschnitte einzeln nach-

einander, um so viele eindeutige URLs wie möglich zu bekommen.

Suchmaschinen-Bots blockieren

Bei Einsatz des Szenarios „Umstellung in einem Durchgang“ oder „Umstellung in Etappen“ und in Abhängigkeit von der Größe der Website und dem Zeitaufwand für die nächsten Schritte kann es sinnvoll sein, die HTTPS-Website bis zum Abschluss der Umstellung für Suchmaschinen-Bots zu blockieren, damit die Suchmaschinen keine widersprüchlichen Signale erhalten. Dazu kann man robots.txt in der HTTPS-Variante einsetzen und dabei festlegen, ob die gesamte HTTPS-Variante oder nur ein Teil davon für das Crawlen gesperrt werden soll. Um alle Bots vollständig zu blockieren, verwenden Sie folgendes Code-Snippet in robots.txt in der HTTPS-Variante:

User-Agent: *

Disallow: /

Hinweis: Dieser Schritt ist optional.

Ob Sie ihn brauchen, hängt hauptsächlich davon ab, wie schnell Ihre Website umgestellt werden kann. Können Sie den Umzug in wenigen Tagen bewerkstelligen, brauchen Sie den Schritt nicht. Das Verfahren kann aber auch verwendet werden, wenn Sie fast alle Aspekte der Umstellung ohne Risiko testen wollen, bevor die Suchmaschinen-Bots den Umzug registrieren.

Im Fall der „Umstellung mithilfe kanonischer URLs“ ist dieser Schritt nicht empfehlenswert.

Ausblick

In der nächsten Ausgabe folgt Teil 2 mit einer detaillierten Beschreibung der technischen Aspekte der eigentlichen Umstellung auf HTTPS im Hinblick auf die Suchmaschinenoptimierung; außerdem erhalten Sie weiterführende wichtige Informationen zur Google Search Console.¶