Engineer

URL canonique

l'adresse de référence d'un contenu, même copié ailleurs

Ce que c’est

Une page accessible à plusieurs adresses peut en désigner une comme son original. L’adresse choisie est son URL canonique : la copie qu’elle demande aux moteurs de recherche et aux autres consommateurs d’indexer, d’afficher et de créditer. Le nom vient du balisage qui la porte, un élément `link` avec `rel="canonical"` . On parle donc aussi de lien canonique, de balise canonique ou simplement de rel=canonical.

```
<link rel="canonical" href="https://example.com/guide/" />
```

Les doublons apparaissent plus souvent qu’on ne l’avait prévu. Un article peut être accessible avec ou sans `www` , en `http` et en `https` , avec un paramètre de suivi à la fin, trié différemment, en vue d’impression ou republié sur un second hôte. La RFC 6596, publiée en 2012, a enregistré cette relation et fixé sa seule règle impérative : la cible doit désigner un contenu qui est soit un doublon de la page référente, soit un sur-ensemble de son contenu. Dans cette limite, la cible peut être la page elle-même, se trouver sur un autre nom d’hôte ou utiliser un autre schéma. La norme HTML la décrit ainsi : « URL préférée pour le document courant » (traduction). Un fichier sans en-tête HTML, comme un PDF, peut envoyer la même déclaration sous forme d’en-tête HTTP `Link` avec `rel="canonical"` .

Une URL canonique est une indication plutôt qu’un ordre. La RFC permet aux applications d’ignorer une désignation incorrecte. La documentation Google classe les redirections et rel=canonical parmi les signaux forts, et l’inclusion dans un sitemap parmi les signaux faibles. Lorsqu’une page ne déclare rien, le moteur de recherche choisit lui-même une copie.

Elle côtoie trois outils avec lesquels on la confond facilement. Une redirection permanente supprime le doublon pour les personnes autant que pour les machines. C'est donc le choix le plus fort chaque fois que l'adresse supplémentaire n'a aucune raison de rester accessible. Une directive `noindex` exclut entièrement une page de l'index, résultat différent de son regroupement avec une autre page. Enfin, `hreflang` relie des pages qui sont des traductions les unes des autres : les traductions sont des équivalents, et aucune n'est un doublon d'une autre.

Utilisez le concept pour prendre une décision

Utilisez « Qu'est-ce qu'une URL canonique ? » pour éclairer une décision réelle. Nous pouvons vous aider à relier ce concept au système que vous construisez ou achetez. Un site ou une application que les assistants d'IA de vos clients peuvent trouver, lire et utiliser : une capacité vérifiée à chaque livraison, plutôt qu'une promesse ponctuelle.

Découvrir nos prestations Contactez-nous →

Vérifiez les hypothèses derrière le terme

Appuyez-vous sur l'explication et les réalisations liées pour formuler les questions de votre projet. Nous pouvons examiner les hypothèses avec vous avant le choix d'un outil. Site Hugo multilingue en 3 langues — 108 fichiers de modèles pour trois langues et quatre formats

Site Hugo multilingue en 3 langues Contactez-nous →

Pourquoi c'est important

L'URL canonique détermine quelle adresse devient celle qui est indexée. Le seul véritable choix consiste à le décider délibérément ou à laisser quelqu'un d'autre le faire. Les moteurs de recherche ne pénalisent pas tant les textes dupliqués qu'ils ne les regroupent : ils gardent une copie et y rattachent les autres. La copie conservée est celle que les personnes voient dans les résultats, à laquelle sont attribués les liens vers chaque variante et à laquelle un robot d'exploration revient. Si on laisse le moteur décider, il peut retenir la version avec un paramètre de campagne, l'hôte de préproduction exposé par inadvertance ou la copie republiée par un partenaire plutôt que l'original.

Choisir entre redirection et URL canonique est surtout une question de personnes. Si personne n'a besoin de la seconde adresse, redirigez-la de manière permanente : le problème disparaît d'un coup pour tous les clients. La RFC 6596 demande elle-même aux auteurs d'écarter la possibilité d'une redirection permanente avant d'utiliser le lien. Si la copie doit rester accessible parce qu'un lien de campagne y mène, qu'une mise en page d'impression y réside ou qu'un second site porte l'article pour son propre public, l'URL canonique est l'outil adapté : elle règle la question pour les index sans déranger les lecteurs.

Le coût tient à la cohérence plutôt qu'à la balise. Une ligne dans l'en-tête coûte peu. Tous les autres emplacements qui indiquent l'adresse de la page doivent donner la même adresse : l'`og:url` des aperçus de liens, les identifiants des données structurées, le sitemap et les liens internes du site. Si deux d'entre eux désignent une adresse et un troisième une autre, la page fait deux déclarations sur elle-même, et un consommateur peut croire l'une ou l'autre. L'URL canonique décide de l'adresse à laquelle sera rattachée toute autorité acquise par le contenu.

À quoi cela ressemble en pratique

Ce site constitue un cas exigeant. Les pages sont publiées sous plusieurs formes à partir d'une même arborescence de contenu. Cette arborescence sert aussi à générer les sous-domaines spécialisés.

Les modèles règlent les doublons par une règle unique : une page n'est son propre original que si elle a été écrite pour l'adresse depuis laquelle elle est servie. Sur le site principal, cela couvre tout : chaque page se désigne donc elle-même. Un sous-domaine spécialisé n'a écrit que sa page d'accueil. Celle-ci se désigne elle-même, tandis que les réalisations en miroir, les pages partagées et chaque liste construite par le générateur pointent vers la copie du site principal au même chemin. La règle de base va dans le sens sûr : tout pointe vers le site principal, sauf déclaration contraire. Une page méritant sa propre identité doit la demander dans ses métadonnées d'en-tête, où la demande est visible. Un nouveau type de page hérite donc de la réponse prudente plutôt que de se déclarer discrètement original.

La valeur est calculée dans un seul modèle, puis lue partout où elle apparaît : par `rel="canonical"`, par `og:url` et `twitter:url`, et par l'`@id` et l'`url` du nœud de données structurées de la page. Le sitemap et les variantes linguistiques posent la même question sous forme de oui ou non. Ils ne sont produits que si l'URL canonique est l'adresse propre de la page : demander l'indexation d'une adresse tout en en nommant une autre comme original est une contradiction qu'un robot peut résoudre en ne faisant confiance à aucune. Une copie sur un hôte spécialisé reste rendue intégralement et son sélecteur de langue fonctionne toujours. Ce qu'elle ne transmet pas, c'est la demande d'indexation, jamais le contenu. La version jumelle Markdown est annoncée depuis la page HTML avec `rel="alternate"` et son type de média. La norme HTML y voit le même document dans un autre format plutôt qu'une page rivale.

Le contrôle de la sortie du site principal exige une seule URL canonique absolue pointant vers la page elle-même sur les pages HTML ordinaires et recherche les barres obliques finales manquantes dans les liens internes vers des pages. Les documents autonomes livrés et la carte sociale suivent leurs propres règles ; les contrôles des sites spécialisés vérifient la distinction entre miroirs et originaux.

Où les problèmes surviennent

Les résultats de recherche montrent une adresse que personne n'a choisie. Un paramètre de suivi, un hôte parasite ou un miroir apparaît comme URL affichée. Soit la page ne déclare aucune URL canonique, soit elle en déclare une tandis que son entrée de sitemap, son `og:url` et ses liens internes en nomment une autre. Le moteur pèse donc des signaux contradictoires et choisit. Produisez une URL canonique absolue par page, y compris une URL pointant vers elle-même pour l'original. Dérivez de la même valeur les adresses du sitemap, des aperçus et des données structurées.

La deuxième page d'une série disparaît de l'index. Chaque page d'une liste paginée ou d'un article découpé désigne la première comme canonique. La première page n'est ni un doublon ni un sur-ensemble de la deuxième. Tout ce qui n'apparaît que dans les pages suivantes peut donc être traité comme doublon et écarté, risque explicitement signalé par la RFC 6596. Laissez chaque page d'une série se désigner elle-même ou faites pointer toutes les pages vers une vue unique contenant réellement l'ensemble du contenu.

Un doublon bloqué dans robots.txt reste affiché, ou l'original disparaît avec lui. Bloquer l'adresse d'un doublon dans robots.txt empêche le robot de lire la page, y compris son URL canonique. Google précise qu'une URL bloquée peut malgré tout être indexée sans son contenu. Marquer les doublons `noindex` les retire entièrement de la recherche au lieu de rattacher ce qu'ils ont acquis à l'original. Laissez les doublons explorables et confiez le regroupement à une URL canonique ou à une redirection.

L'URL canonique déclarée est ignorée. La cible redirige de manière permanente ailleurs, déclare elle-même une autre URL canonique, renvoie 404 ou entre en concurrence avec un second élément canonique sur la même page. Un script peut aussi réécrire le lien après l'arrivée du HTML. Chacun de ces cas rend la désignation incorrecte, et la RFC permet à un consommateur de revenir à son propre jugement. Pointez vers l'adresse finale renvoyant 200, produisez un seul élément dans l'en-tête HTML servi et ne laissez aucun script le modifier ensuite.

Pour aller plus loin

- RFC 6596 : la relation de lien canonique — l'enregistrement de la relation, la règle du doublon ou du sur-ensemble et les désignations à éviter.
- Norme HTML : type de lien « canonical » — la définition du mot-clé en HTML et les endroits où il peut apparaître.
- RFC 8288 : liens Web — l'en-tête HTTP `Link` portant une URL canonique pour les fichiers dépourvus d'en-tête HTML.
- Google Search Central : regrouper les URL en double — la manière dont un moteur de recherche pondère redirections, URL canoniques et sitemaps les uns par rapport aux autres.

Glossaire

Choisissez selon vos contraintes réelles

Présentez le processus de travail, les interfaces existantes et les utilisateurs qui ont besoin du changement. Nous pouvons définir une livraison testable et une transmission permettant à votre équipe d'assurer la maintenance des éléments livrés.

Consulter les tarifs Contactez-nous →

Rendez la terminologie concrète

Demandez comment « Qu'est-ce qu'une URL canonique ? » s'applique à votre organisation. Un exemple concret tiré de votre système actuel suffit pour engager la discussion.

Découvrir nos prestations

welcome@engineer.company +45 50 10 61 84 Copenhague, Danemark

Emportez cette page avec vous

Fiche de contact

Créé avec soin et amour par D.

CVR 45969703 · Tagensvej 105C, 3. th, 2200 København N

Droit d'auteur © 2025 – aujourd'hui · Engineer ApS · Tous droits réservés.

Imprimé depuis https://engineer.company/fr/glossary/canonical/