Engineer

正規URL

検索エンジンなどに正本として扱うよう求めるページのURL

どのようなものか

複数のアドレスで応答するページは、その一つを原本として指定できます。指定したアドレスが正規URLです。検索エンジンなど、ページを利用する側に、インデックスへ登録し、表示し、評価を集約してほしいコピーを示します。名前は、その指定を記述する、`rel="canonical"` 付きの `link` 要素に由来します。同じものをcanonical link（正規リンク）、canonical tag（正規タグ）、または単にrel=canonicalとも呼びます。

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

重複は、意図した以上によく生まれます。一つの記事が、`www` の有無、`http` と `https` 、末尾の追跡パラメーター、異なる並べ方、印刷用の表示、別のホストでの再公開など、複数の形で開けます。2012年に公開されたRFC 6596は、このリンク関係を登録し、一つの厳格なルールを定めました。リンク先は、指定元のページと同じ内容か、その内容をすべて含むものでなければなりません。その範囲なら、ページ自身でも、別のホスト名でも、別のスキームでも指定できます。HTML標準は、「現在の文書を表す優先URL」と説明しています。PDFのようにHTMLのheadを持たないファイルも、`rel="canonical"` 付きのHTTP `Link` ヘッダーで、同じ指定を送れます。

正規URLの指定は、命令ではなく手掛かりです。RFCは、不適切な指定をアプリケーションが無視することを認めています。Googleの文書は、リダイレクトとrel=canonicalを強い信号、サイトマップへの掲載を弱い信号としています。ページが何も指定しなければ、検索エンジンが自らコピーを選びます。

混同しやすい、近い三つの仕組みがあります。恒久的なリダイレクトは、機械だけでなく、人に対しても重複をなくすため、余分なアドレスを開けるままにしておく理由がなければ、より確実な選択です。 `noindex` ディレクティブは、別のペ

ージへまとめるのではなく、そのページをインデックスから完全に除きます。`hreflang`は、互いの翻訳であるページを結びます。翻訳は対応するページであり、どれも他のページの重複ではありません。

概念を判断に生かす

「正規URLとは？」を実際の判断に生かしましょう。その概念を、開発または購入するシステムに結び付けて考えるお手伝いをします。 お客様のAIアシスタントがサイトやアプリを見つけ、読み取り、操作できる状態を、最初の約束だけで済ませず、リリースごとに検証します。

サービスを見る お問い合わせ →

用語の背後にある前提を検証する

解説とリンク先の実績を参考に、プロジェクトについての質問を整理してください。ツールを選ぶ前に、前提を一緒に検討できます。 3言語の多言語Hugoサイトを構築 — 108 3言語・4形式を提供するテンプレートファイル

3言語の多言語Hugoサイトを構築 お問い合わせ →

なぜ重要なのか

正規URLが決めるのは、どのアドレスをインデックスへ登録するかです。実際の選択肢は、意識的に決めるか、他の誰かに決められるかしかありません。検索エンジンは、重複した文章を罰するというより、一つにまとめます。一つのコピーを残し、他をそこへ集約します。残したコピーが検索結果に表示され、各版へのリンクから得る評価が集まり、クローラーが再訪する対象になります。エンジンに任せると、キャンペーンのパラメーター付きの版、意図せず公開されたステージング用ホスト、原本でなく提携先の再公開版が選ばれる可能性があります。

リダイレクトと正規URLのどちらを使うかは、主に、人にとっての問題です。二つ目のアドレスが誰にも必要なければ、恒久的にリダイレクトすれば、すべてのクライアントで問題が一度に消えます。RFC 6596自体も、リンクを使う前に、恒久的なリダイレクトを使えないか確認するよう求めています。キャンペーンのリンクがそこを指す、印刷用レイアウトがある、別のサイトが自分の読者向けに記事を載せるなど、コピーを開ける状態にしておく必要があれば、正規URLを使います。インデックス向けの問いを解決し、読者の利用には手を加えないためです。

負担がかかるのは、タグではなく、一貫性です。head内の一行を書く手間は小さくても、ページのアドレスを示す他のすべての場所を、それと一致させる必要があります。リンクプレビューの`og:url`、構造化データの識別子、サイトマップ、サイト内部のリンクです。二つが一つのアドレス、三つ目が別のアドレスを示すと、ページが自分について二つの主張をすることになり、利用する側がどちらを採用するかは、分かりません。コンテンツが得る評価をどのアドレスに結び付けるかが、正規URLの決めることです。

実際の活用例

このサイトは、正規URLが必要となる複雑な事例です。同じコンテンツのツリーから、各ページを複数の表現形式で公開しています。そのツリーから、専門分野別のサブドメイン向けにもビルドします。

テンプレートは、一つのルールで重複を整理します。その配信先のアドレスに向けて書いたページだけを、自分自身が原本であると扱います。主サイトではすべてが該当するため、各ページが自分を指定します。専門分野別のサブドメイン向けに独自に書かれたのはトップページだけなので、それは自分を指定しますが、複製した実績、共通ページ、生成処理が作るすべての一覧は、同じパスの主サイトのコピーを指定します。既定の動作は安全な方向です。別の指定がなければ、すべて主サイトを指し、独自の正本として扱うべきページは、front matterでその扱いを明示的に指定する必要があります。そこで指定すれば、判断が見えます。そのため、新しい種類のページにも、黙って原本を名乗らせず、安全側の判断を引き継がせられます。

値は一つのテンプレートで計算し、現れるすべての場所で使います。`rel="canonical"`、`og:url`、`twitter:url`、ページの構造化データのノードの`@id`と`url`です。サイトマップと他言語版への参照も、同じ条件を満たすかどうかを判定し、正規URLがページ自身のアドレスである場合だけ出力します。あるアドレスをインデックスへ登録するよう求めながら、原本には別のアドレスを指定するのは矛盾であり、クローラーは両方を信用しないことで解決できてしまうためです。専門分野別のホストのコピーも、全文を描画し、言語切り替えも動きます。出さないのはインデックスへの登録要求であり、内容ではありません。Markdown版は、HTMLページから、`rel="alternate"`とメディアタイプを付けて知らせます。HTML標準は、これを競合するページではなく、同じ文書の別形式と解釈します。

主サイトの出力検査では、通常のHTMLページに、自分自身を指す絶対URLの正規URLが一つあることを求め、内部ページへのリンクで末尾のスラッシュが欠けていないかを調べます。別ホストへ配信する独立した文書とソーシャルカードには、それぞれの規則があります。専門分野別ホストの検査では、ミラーと原本の区別を確かめます。

問題が起きるところ

誰も選んでいないアドレスが、検索結果に出ます。 追跡パラメーター、意図しないホスト、ミラーが、表示URLになります。ページが正規URLを指定していないか、指定しても、サイトマップ、`og:url`、内部リンクが別のアドレスを示し

ているため、エンジンは混ざった信号を比較して選びます。すべてのページに絶対URLの正規URLを一つ出力し、原本にも自分自身を指すものを置きます。サイトマップ、プレビュー、構造化データのアドレスも、同じ値から導きます。

ページ分割した二ページ目が、インデックスから消えます。 ページ分割した一覧や記事で、全ページが一ページ目を正規URLにしています。一ページ目は二ページ目と同一でも、その内容をすべて含むものでもないため、後続だけにある内容が、重複として無視される場合があります。RFC 6596は、この問題を名指しで警告しています。各ページが自分を指定するか、本当に全内容を含む一つの全件表示ページへ、すべてを向けます。

robots.txtで拒否した重複が、掲載されたままになったり、原本まで消えたりします。 robots.txtで重複アドレスを拒否すると、クローラーは、正規URLの指定も含め、ページを読めません。Googleは、拒否したURLでも、内容なしでインデックスへ登録され得るとしています。重複に `noindex` を付けると、得た評価を原本へ集めるのではなく、検索から完全に除きます。重複もクロール可能にして、正規URLかリダイレクトに集約を任せます。

指定した正規URLが無視されます。 リンク先が別の場所へ恒久的にリダイレクトする、さらに別の正規URLを指定する、404を返す、同じページの二つ目の正規URL要素と競合するといった場合です。また、HTMLが届いた後に、スクリプトがリンクを書き換えることもあります。いずれも指定を不適切にするため、RFCは、利用する側が自分の判断へ戻ることを認めます。200を返す最終アドレスを指し、配信するHTMLのheadに要素を一つだけ出力し、その後はスクリプトで変更させないようにします。

参考資料

- RFC 6596：The Canonical Link Relation：リンク関係の登録、重複または全内容を含むというルール、避けるべき指定を説明します。
- HTML標準：リンク種別canonical：HTMLでのキーワードの定義と、記述できる場所を説明します。
- RFC 8288：Web Linking：HTMLのheadを持たないファイルで正規URLを伝える、HTTP `Link` ヘッダーを説明します。
- Google Search Central：重複URLの集約：ある検索エンジンが、リダイレクト、正規URL、サイトマップをどのように比較するかを説明します。

用語集

実際の制約に合わせて選ぶ

業務の流れ、既存のインターフェース、変更を必要とする利用者をお知らせください。検証可能な納品内容と、お客様のチームが保守を続けられる引き継ぎ内容を定められます。

料金を見る お問い合わせ →

用語を実務に結び付ける

「正規URLとは？」がお客様の組織にどう当てはまるかをご相談ください。現在のシステムの具体例が一つあれば、話を始められます。

サービスを見る

welcome@engineer.company +45 50 10 61 84 デンマーク、コペンハーゲン

このページを持ち歩く

連絡先カード

心を込めて丁寧に制作： D.

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

著作権 © 2025 – 現在 · Engineer ApS · すべての権利を留保します。

印刷元： https://engineer.company/ja/glossary/canonical/