Engineer

Канонічна URL

єдина адреса, яку сторінка називає своєю

Що це таке

Сторінка, доступна за кількома адресами, може назвати одну з них своїм оригіналом, і названа адреса — це її канонічна URL: та копія, яку сторінка просить пошукові системи та інших споживачів індексувати, показувати й зараховувати їй заслуги. Назва походить від розмітки, що її передає, елемента `link` з `rel="canonical"`, тому те саме називають ще канонічним посиланням, тегом canonical або просто rel=canonical.

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

Дублікати з'являються частіше, ніж хтось планує. Ту саму статтю можна відкрити з `www` і без нього, через `http` і `https`, з параметром відстеження в кінці, з іншим сортуванням, у версії для друку або в копії на іншому домені. RFC 6596, опублікований 2012 року, зареєстрував це відношення і встановив для нього одне тверде правило: ціль мусить указувати на вміст, який є або дублікатом сторінки, що посилається, або її надмножиною. У цих межах ціль може бути самою сторінкою, лежати на іншому імені хоста чи використовувати інший протокол. Стандарт HTML описує її як бажану URL поточного документа. Файл без елемента `head`, як-от PDF, може передати те саме твердження заголовком HTTP `Link` з `rel="canonical"`.

Канонічна URL — підказка, а не наказ. RFC дозволяє програмам знехтувати хибним призначенням, а документація Google вважає переспрямування й rel=canonical сильними сигналами, а наявність у карті сайту — слабким. Якщо сторінка нічого не оголошує, пошукова система сама обирає копію.

Поряд із нею стоять три інструменти, які легко з нею сплутати. Постійне переспрямування прибирає дублікат і для людей, і для машин, тому воно сильніше, коли зайва адреса нікому не потрібна. Директива `noindex` тримає сторінку поза індексом узагалі, а це інший результат, ніж злиття з іншою сторінкою. А `hreflang` пов’язує сторінки, які є перекладами одна одної; переклади рівноцінні, і жоден із них не є дублікатом іншого.

Застосуйте поняття до рішення

Застосуйте «Що таке канонічна URL?» до конкретного рішення. Ми можемо пов'язати поняття із системою, яку ви створюєте або купуєте. Сайт чи застосунок, який ШІ-асистенти ваших клієнтів можуть знайти, прочитати й використати, і це перевіряється з кожним випуском, а не обіцяється один раз.

Переглянути наші послуги Зв'язатися з нами →

Перевірте припущення за терміном

Використайте пояснення та пов'язані роботи, щоб сформулювати запитання для проєкту. Ми можемо перевірити припущення разом до вибору інструмента. Побудувала сайт на Hugo 3 мовами — 108 файлів шаблонів обслуговують три мови й чотири формати

Побудувала сайт на Hugo 3 мовами Зв'язатися з нами →

Чому це важливо

Канонічна URL вирішує, яка адреса стане індексованою, і справжній вибір лише в тому, чи ухвалити це рішення свідомо, чи дати ухвалити його комусь іншому. Пошукові системи не стільки карають за дубльований текст, скільки згортають його: залишають одну копію, а решту вливають у неї. Залишена копія — це та, яку люди бачать у результатах пошуку, та, якій зараховують посилання на всі варіанти, і та, до якої повертається пошуковий робот. Якщо віддати вибір пошуковій системі, нею може стати версія з параметром кампанії, тестове середовище, що випадково потрапило назовні, або передрук у партнера замість оригіналу.

Переспрямування чи канонічна URL — це здебільшого питання про людей. Якщо другою адресою ніхто не користується, переспрямуйте її постійно, і проблема зникне для всіх клієнтів одразу; RFC 6596 і сам радить спершу перевірити, чи не впорається з завданням переспрямування. Якщо копія мусить лишатися доступною, бо на неї веде посилання з кампанії, там живе версія для друку або інший сайт публікує статтю для своєї аудиторії, тоді інструмент — канонічна URL: вона вирішує питання для індексів і не заважає читачам.

Ціна полягає в узгодженості, а не в самому тегу. Один рядок у `head` нічого не коштує. Але кожне інше місце, що називає адресу сторінки, мусить із ним збігатися: `og:url`, який використовують попередні перегляди посилань, ідентифікатори у структурованих даних, карта сайту і внутрішні посилання самого сайту. Якщо два з них називають одну адресу, а третє іншу, сторінка робить про себе два різні твердження, і споживач має право повірити будь-якому. Канонічна URL вирішує, до якої адреси прив'язується вага, яку здобуде вміст.

Як це виглядає на практиці

Цей сайт — складний випадок. З одного дерева вмісту публікують сторінки в кількох поданнях. Це саме дерево також збирають для тематичних субдоменів.

Шаблони розв'язують питання дублікатів одним правилом: сторінка є власним оригіналом лише тоді, коли її написано для тієї адреси, з якої її віддають. На основному сайті це стосується всього, тож кожна сторінка називає саму себе. Тематичний субдомен написав свою головну сторінку і більше нічого, тож ця сторінка називає себе, а дзеркальні роботи, спільні сторінки й усі списки, які будує генератор, указують на копію основного сайту за тим самим шляхом. Типова поведінка працює в безпечний бік: усе вказує додому, якщо не сказано інакше, а сторінка, що заслуговує на власну ідентичність, мусить попросити про неї у front matter, де цей запит видно. Тож новий вид сторінки успадковує обережну відповідь і не видає себе тихцем за оригінал.

Значення обчислюється в одному шаблоні й читається всюди, де воно з'являється: у `rel="canonical"`, в `og:url` і `twitter:url`, а також в `@id` і `url` вузла структурованих даних сторінки. Карта сайту й мовні альтернативи ставлять те саме питання у формі «так чи ні» і з'являються лише тоді, коли канонічна URL збігається з власною адресою сторінки, адже просити проіндексувати одну адресу й водночас називати оригіналом іншу — це суперечність, яку пошуковий робот може розв'язати, не повіривши жодній. Копія на тематичному субдомені все одно показується повністю, і перемикач мов працює; вона притримує запит на індексацію, але ніколи не вміст. Двійника в Markdown HTML-сторінка оголошує через `rel="alternate"` з його типом медіа, і стандарт HTML читає це як той самий документ в іншому форматі, а не як сторінку-суперницю.

Перевірка основного сайту вимагає одну абсолютну канонічну URL, що вказує на саму сторінку, для звичайних HTML-сторінок і шукає відсутні кінцеві скісні риски у внутрішніх посиланнях. Окремі документи для іншого хоста й соціальна картка мають власні правила; перевірки тематичних хостів розрізняють дзеркала й оригінали.

Де виникають проблеми

Результати пошуку показують адресу, яку ніхто не обирав. Параметр відстеження, чуже ім'я хоста чи дзеркало з'являються як показана URL. Сторінка або не оголошує канонічної URL, або оголошує її, тоді як запис у карті сайту, `og:url` і внутрішні посилання називають щось інше, тож пошукова система зважує змішані сигнали й обирає сама. Ставте одну абсолютну канонічну URL на кожну сторінку, на оригіналі теж (там вона вказує на саму сторінку), і виводьте адреси для карти сайту, попереднього перегляду та структурованих даних із цього самого значення.

Друга сторінка серії випадає з індексу. Кожна сторінка списку з пагінацією чи розділеної статті називає канонічною першу сторінку. Перша сторінка не є ні дублікатом, ні надмножиною другої, тож усе, що є лише на наступних сторінках, можуть визнати дублікатом і відкинути, і RFC 6596 прямо застерігає від цього. Нехай кожна сторінка серії називає себе або всі вони вказують на одну сторінку з повним текстом, яка справді містить усе.

Дублікат, закритий у robots.txt, лишається в результатах, або оригінал зникає разом із ним. Закрита в robots.txt адреса дубліката не дає пошуковому роботу прочитати сторінку, а отже і її канонічну URL, і Google зазначає, що закриту URL усе одно можуть проіндексувати без вмісту. Позначка `noindex` на дублікатах прибирає їх із пошуку повністю замість того, щоб передати здобуте ними оригіналу. Залишайте дублікати доступними для пошукового робота, а злиття доручіть канонічній URL або переспрямуванню.

Оголошену канонічну URL ігнорують. Ціль постійно переспрямовує далі, сама називає іншу канонічну URL, відповідає кодом 404 або конкурує з другим елементом canonical на тій самій сторінці; ще скрипт може переписати посилання вже після того, як надійшов HTML. Кожен із цих випадків робить призначення хибним, і RFC дозволяє споживачеві

покластися на власне судження. Указуйте на кінцеву адресу, що відповідає кодом 200, ставте один елемент у `head` того HTML, який віддає сервер, і не дозволяйте жодному скрипту змінювати його згодом.

Що почитати далі

- RFC 6596: The Canonical Link Relation — реєстрація відношення, правило «дублікат або надмножина» і призначення, яких слід уникати.
- HTML Standard: link type “canonical” — як HTML визначає це ключове слово і де воно може стояти.
- RFC 8288: Web Linking — заголовок HTTP `Link`, що передає канонічну URL для файлів без елемента `head`.
- Google Search Central: consolidate duplicate URLs — як одна пошукова система зважує переспрямування, канонічні URL і карти сайту одне проти одного.

Словник

Обирайте за фактичними умовами

Опишіть робочий процес, наявні інтерфейси та користувачів, яким потрібна зміна. Ми можемо визначити перевірюваний результат і спосіб передачі, який дасть вашій команді змогу підтримувати його.

Дивитися, скільки це коштує Зв'язатися з нами →

Зробіть термінологію практичною

Запитайте, як «Що таке канонічна URL?» стосується вашої організації. Конкретного прикладу з поточної системи достатньо для початку розмови.

Переглянути наші послуги

welcome@engineer.company +45 50 10 61 84 Копенгаген, Данія

Візьміть цю сторінку із собою

Візитівка

Дбайливо зроблено з любов'ю — D.

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

Авторське право © 2025 – дотепер · Engineer ApS · Усі права захищено.

Надруковано з https://engineer.company/uk/glossary/canonical/