Engineer

Карта сайту

перелік сторінок сайту, написаний для машин

Що це таке

Карта сайту — це файл, який вебсайт публікує сам про себе: адреси, які власник хоче показати пошуковим системам, по одному запису на сторінку, кожен із кількома необов'язковими відомостями. Вона потрібна тому, що інакше краулер пізнає сайт повільно, переходячи за посиланнями зі сторінок, які вже знає, а деякі сторінки стоять у кінці довгого ланцюжка посилань або майже не мають посилань на себе.

Формат, спільний для великих пошукових систем, — протокол Sitemaps, опублікований на sitemaps.org. Його XML-версія загортає перелік в елемент `<urlset>` з одним `<url>` на сторінку. Усередині `<loc>` містить повну адресу і є єдиним обов'язковим полем. `<lastmod>` — дата останньої зміни сторінки, `<changefreq>` — припущення, як часто вона змінюватиметься, а `<priority>` — значення від 0,0 до 1,0, що ставить сторінку в ряд з іншими сторінками того самого сайту і ніколи — з чужими. Протокол приймає також простий текстовий файл з однією адресою на рядок або стрічку RSS чи Atom.

Протокол має жорсткі межі. Один файл містить щонайбільше 50 000 адрес і 50 МБ без стиснення, його можна стиснути gzip, він має бути в UTF-8 з екранованими значеннями, а кожна адреса — коротша за 2 048 символів. Більший сайт ділить перелік і публікує індекс карт сайту — файл іншого типу, що називає до 50 000 карт на тому самому сайті. Крім того, карта може містити лише адреси на рівні свого розташування або нижче: файл `/catalog/sitemap.xml` охоплює тільки сторінки всередині `/catalog/`.

Пошукова система дізнається, де лежить файл, одним із двох способів. Рядок `Sitemap:` у robots.txt дає його повну адресу незалежно від правил для окремих краулерів, або власник надсилає файл через інструменти самої пошукової системи, наприклад Google Search Console.

Так само називають і сторінку з посиланнями для людей; коли різниця важлива, ту, що написана для програм, називають «XML-картою сайту».

Застосуйте поняття до рішення

Застосуйте «Що таке карта сайту?» 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

Документація Google відверто говорить про межі. Сайт, сторінки якого добре пов’язані посиланнями, зазвичай можна знайти й без карти сайту, а межу «малого» сайту Google проводить приблизно на 500 сторінках, які мають з’являтися в результатах. Google також пише, що карта сайту не гарантує сканування й індексування всього, що в ній є, і називає її надсилання лише підказкою. Найбільше користі вона дає великим сайтам, новим сайтам із малою кількістю зовнішніх посилань і сайтам, де багато відео, зображень чи новин.

Майже кожен генератор сайтів створює карту сайту безкоштовно, тож справжнє рішення — що саме стверджує файл, а кожен запис є твердженням, яке пошукова система може звірити з самою сторінкою.

Які адреси туди потрапляють? Google просить канонічну адресу кожної сторінки. Перелік, роздутий переспрямуваннями, дублікатами чи сторінками з позначкою `noindex`, просить саме того, що решта сайту велить оминути.

Які поля? Google каже, що ігнорує `<priority>` і `<changefreq>`. `<lastmod>` він використовує, лише якщо дата стабільно точна і це можна перевірити, наприклад порівнявши її з останньою зміною сторінки, а протокол вимагає, щоб дата відповідала дню останньої зміни сторінки, а не дню створення карти. Істотною Google вважає зміну основного вмісту, структурованих даних чи посилань, а новий рік у рядку про авторські права — дрібницею.

Тут і криється ціна. Дата з файлової системи нічого не коштує й зазвичай хибна, бо розгортання, експорти та свіжі клони репозиторію торкаються файлів, текст яких не змінився. Дата з історії версій чесна, але тоді збірці потрібна ця історія під рукою і правильний порядок її читання. Дата, введена вручну, точна, доки хтось не забуде її оновити.

Як це виглядає на практиці

Файл robots.txt цього сайту називає одну адресу, `https://engineer.company/sitemap.xml`, і цей файл є індексом, що перелічує карти сайту для опублікованих мовних версій. Кожен запис про сторінку містить адресу й дату, а також посилання на ту саму сторінку в інших її опублікованих перекладах. `<priority>` і `<changefreq>` пропущено: Google каже, що ігнорує обидва, а протокол не вимагає жодного з них. Коли писалася ця стаття, запис для сторінки про готовність до ШІ-агентів виглядав так, без мовних посилань:

```
<url>
  <loc>https://engineer.company/uk/ai-ready/</loc>
  <lastmod>2026-10-15T14:25:16+02:00</lastmod>
</url>
```

Перелік вибірковий навмисно. Тематичні сайти-супутники компанії перевидають роботи звідси з канонічною адресою, що вказує назад, і той самий шаблон не пускає ці копії в карту сайту супутника: запис просить проіндексувати саме цю адресу, а просити про копію й водночас називати оригіналом іншу сторінку означало б суперечити собі. Кожен хост оголошує лише власні сторінки.

Дати мають власну історію. Свого часу карта сайту казала пошуковим системам, що 172 зі 176 сторінок змінилися тієї самої миті. Причина була буденна: дата бралася з часу зміни кожного файлу, а експорт вмісту переписує кожен файл за кожного прогону. Виправлення перенесло джерело до історії версій, із ланцюжком запасних варіантів, що пробує спершу явне поле, потім історію комітів, потім файл. Одна пастка варта уваги: буквальна назва поля має бути в переліку, інакше генератор ніколи його не прочитає. Після цього повторна синхронізація після місяця редагування торкнулася восьми файлів зі 186 замість усіх, і карта сайту каже щось правдиве.

Тепер ця дата працює не лише в карті сайту. Як описує сторінка про готовність до ШІ-агентів, кожне розгортання оголошує зміни через IndexNow — протокол, у якому беруть участь Bing і ще кілька пошукових систем. Скрипт розгортання визначає, що змінилося, порівнюючи `lastmod` кожної сторінки з тим, що надіслав минулого разу, тож розгортання, яке змінює одну сторінку, оголошує одну сторінку. Google немає у власному переліку учасників IndexNow, тож про зміну він дізнається, коли наступного разу прочитає карту сайту. Та сама сторінка описує запобіжник: збірка звіряє кожну дату в карті сайту з історією сторінки і не публікує сайт, якщо щось не так. Для людей, а не програм, є ще зручний для читання покажчик усіх сторінок.

Де виникають проблеми

Кожна сторінка твердить, що змінилася сьогодні. Карта сайту має одну дату, або одну на кожне розгортання, для сотень сторінок, і пошукова система, що її читає, не дізнається, яка робота нова. Причина — дата з файлової системи чи з годинника збірки. Google використовує `<lastmod>`, лише коли дата стабільно точна і це можна перевірити, а спільна позначка часу такою явно не є. Беріть дату з власної історії вмісту і після кожної зміни збірки звіряйте з нею вибірку записів.

Перелік просить сторінки, які сайт хоче лишити поза пошуком. Переспрямування, дублікати з параметрами відстеження, сторінки з `noindex`, тестові хости й копії, чия канонічна адреса вказує деінде, потрапляють у карту сайту від генератора, що вносить кожен зібраний файл. Кожна з них відправляє краулер на адресу, від якої сама сторінка потім відмовляється. Вносьте лише канонічну адресу кожної сторінки, призначеної для результатів, і нехай карта сайту та тег canonical відповідають за одним правилом, як це роблять шаблони цього сайту.

Записи виходять за межі, які файл може охоплювати. Карта сайту в `/blog/sitemap.xml` містить сторінки магазину або записує відносні шляхи на кшталт `/page.html`. Протокол обмежує файл адресами на рівні його власної теки або нижче, а Google сканує адреси точно так, як їх записано. Публікуйте карту в корені, записуйте кожен `<loc>` як повну адресу `https://` на тому самому хості й переходьте на індекс карт, перш ніж перелік наблизиться до 50 000 записів або 50 МБ.

Сторінки місяцями стоять у карті, але в індекс не потрапляють. Власник сприймає карту сайту як наказ і чекає. Google називає її підказкою: адресу з переліку однаково можуть пропустити або злити з дублікатом із причин, що лежать поза файлом. Дивіться на сторінку, а не на перелік: чи посилаються на неї інші сторінки, чи вказує її canonical на неї саму і чи не тримають краулер осторонь robots.txt або правило `noindex` .

Що почитати далі

- Sitemaps XML format — сам протокол: кожен тег, обмеження розміру, індексні файли і які адреси може містити файл.
- Build and submit a sitemap — правила Google, зокрема теги, які він ігнорує, і коли він довіряє даті.
- Learn about sitemaps — коли, на думку Google, сайту потрібна карта, а коли досить добрих посилань.
- IndexNow documentation — протокол, що повідомляє пошукові системи-учасниці про зміну в день її публікації.
- RFC 9309: Robots Exclusion Protocol — стандарт robots.txt, який дозволяє краулеру прочитати рядок `Sitemap:` поруч із правилами.

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 карта сайту?» 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/sitemap/