Engineer

Sitemap

Liste der Seiten einer Website für Maschinen

Was es ist

Eine Sitemap ist eine Datei, die eine Website über sich selbst veröffentlicht: die Adressen, die ihr Eigentümer Suchmaschinen bekannt machen möchte, mit einem Eintrag pro Seite und einigen optionalen Angaben dazu. Ohne sie lernt ein Crawler eine Website auf dem langsamen Weg kennen, indem er Links von schon bekannten Seiten folgt. Manche Seiten stehen aber am Ende einer langen Linkkette oder haben kaum eingehende Links.

Das gemeinsame Format der großen Suchmaschinen ist das unter sitemaps.org veröffentlichte Sitemaps-Protokoll. Seine XML-Fassung umschließt die Liste mit einem Element `<urlset>` und darin einem `<url>` pro Seite. Dessen `<loc>` enthält die vollständige Adresse und ist das einzige verpflichtende Kindelement. `<lastmod>` nennt das letzte Änderungsdatum der Seite, `<changefreq>` schätzt ihre künftige Änderungshäufigkeit. `<priority>` ist ein Wert von 0,0 bis 1,0, der die Seite nur gegenüber den übrigen Seiten derselben Website einordnet, niemals gegenüber fremden. Das Protokoll akzeptiert auch eine reine Textdatei mit einer Adresse pro Zeile oder einen RSS- beziehungsweise Atom-Feed.

Das Protokoll setzt feste Grenzen. Eine Datei enthält höchstens 50.000 Adressen und 50 MB unkomprimierte Daten, darf mit gzip komprimiert sein und muss UTF-8 verwenden. Sonderzeichen in ihren Werten müssen als Entitäten maskiert sein. Jede Adresse bleibt unter 2.048 Zeichen. Eine größere Website teilt ihre Liste auf und veröffentlicht einen Sitemap-Index. Dieser zweite Dateityp benennt bis zu 50.000 Sitemaps derselben Website. Eine Sitemap darf außerdem nur Adressen an oder unter ihrem eigenen Speicherort auflisten: Eine Datei unter `/catalog/sitemap.xml` deckt die Seiten unter `/catalog/` ab und endet an dieser Ordnergrenze.

Eine Suchmaschine erfährt den Speicherort auf einem von zwei Wegen. Eine Zeile `Sitemap:` in robots.txt nennt die vollständige Adresse unabhängig von jeder User-Agent-Gruppe. Oder der Eigentümer reicht die Datei über die Werkzeuge der Suchmaschine ein, etwa die Google Search Console.

Dasselbe Wort bezeichnet auch eine Linkseite für menschliche Besucher. Wenn die Unterscheidung wichtig ist, heißt die für Programme bestimmte Datei „XML-Sitemap“.

Das Konzept für eine Entscheidung nutzen

Nutzen Sie „Was ist eine Sitemap?“ für eine konkrete Entscheidung. Wir können helfen, das Konzept mit dem System in Verbindung zu bringen, das Sie entwickeln oder einkaufen. Eine Website oder App, die die KI-Assistenten Ihrer Kunden finden, lesen und bedienen können — bei jeder Veröffentlichung geprüft statt einmal versprochen.

Unsere Leistungen entdecken Kontakt aufnehmen →

Die Annahmen hinter dem Begriff prüfen

Nutzen Sie die Erklärung und die verlinkten Arbeiten, um Fragen für Ihr Projekt zu formulieren. Wir können die Annahmen gemeinsam mit Ihnen prüfen, bevor ein Werkzeug ausgewählt wird.
Hugo-Website mehrsprachig in 3 Sprachen aufgebaut — 108 Vorlagendateien für drei Sprachen und vier Formate

Hugo-Website mehrsprachig in 3 Sprachen aufgebaut Kontakt aufnehmen →

Warum es wichtig ist

Googles Dokumentation benennt die Grenzen offen. Eine Website mit gut verlinkten Seiten lässt sich meist ohne Sitemap entdecken. Die Grenze für eine „kleine“ Website zieht Google bei ungefähr 500 Seiten, die in Suchergebnissen erscheinen sollen. Eine Sitemap „garantiert nicht, dass alle darin enthaltenen Einträge gecrawlt und indexiert werden“, und ihre Einreichung „ist lediglich ein Hinweis“. Am meisten hilft sie großen Websites, neuen Websites mit wenigen Links von außen und Websites mit vielen Videos, Bildern oder Nachrichten.

Fast jeder Site-Generator erzeugt eine kostenlos. Die eigentliche Entscheidung ist deshalb, was die Datei behauptet. Jeder Eintrag ist eine Aussage, die eine Suchmaschine an der Seite selbst überprüfen kann.

Welche Adressen gehören hinein? Google verlangt die kanonische Adresse jeder Seite. Eine mit Weiterleitungen, Duplikaten oder mit `noindex` markierten Seiten aufgefüllte Liste fordert genau das an, was die übrige Website ausschließt.

Welche Felder? Google ignoriert nach eigener Aussage `<priority>` und `<changefreq>`. `<lastmod>` verwendet es nur, „wenn es durchgängig und überprüfbar korrekt ist, etwa durch einen Vergleich mit der letzten Änderung der Seite“. Das Protokoll verlangt, dass das Datum „auf die letzte Änderung der verlinkten Seite gesetzt werden muss, nicht auf die Erzeugung der Sitemap“. Google wertet Änderungen am Hauptinhalt, an strukturierten Daten oder Links als wesentlich, ein neues Copyright-Jahr dagegen als unbedeutend.

Darin steckt der Aufwand. Ein Datum aus dem Dateisystem ist kostenlos und meist falsch, weil Deployments, Exporte und frische Checkouts Dateien berühren, deren Wortlaut unverändert blieb. Ein Datum aus dem Versionsverlauf ist ehrlich, aber der Build muss diesen Verlauf vorliegen haben und in der richtigen Reihenfolge lesen. Ein von Hand eingetragenes Datum bleibt genau, bis jemand seine Änderung vergisst.

Wie es in der Praxis aussieht

Die robots.txt dieser Website nennt eine Adresse, `https://engineer.company/sitemap.xml`. Diese Datei ist ein Sitemap-Index, der die Sitemaps der veröffentlichten Sprachfassungen auflistet. Jeder Seiteneintrag enthält die Adresse und ein Datum sowie Links zur gleichen Seite in ihren anderen veröffentlichten Übersetzungen. `<priority>` und `<changefreq>` fehlen, weil Google nach eigener Aussage beide ignoriert und das Protokoll keines der beiden Elemente verlangt. Als dieser Artikel entstand, sah der Eintrag für die Seite zur KI-Bereitschaft ohne die Sprachlinks so aus:

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

Die Liste enthält bewusst nur eine Auswahl. Die thematisch ausgerichteten Satelliten-Websites des Unternehmens veröffentlichen Arbeiten von hier erneut, mit einem Canonical-Verweis zurück zum Original. Dieselbe Vorlage hält diese Kopien aus der Sitemap jedes Satelliten heraus: Ein Eintrag bittet eine Suchmaschine, genau diese Adresse zu indexieren. Eine Kopie anzufordern und zugleich eine andere Seite als Original zu benennen, wäre widersprüchlich. Jeder Host meldet nur seine eigenen Seiten.

Die Datumsangaben haben eine eigene Geschichte. Zu einem Zeitpunkt meldete die Sitemap Suchmaschinen, dass sich 172 ihrer 176 Seiten im selben Augenblick geändert hätten. Die Ursache war alltäglich: Das Datum kam aus der Änderungszeit jeder Datei, und der Inhaltsexport schreibt bei jedem Lauf jede Datei neu. Die Korrektur verlegte die Quelle in den Versionsverlauf, mit einer Fallback-Kette, die zuerst ein explizites Feld, dann den Commit-Verlauf und zuletzt die Datei prüft. Darin lag eine wissenswerte Falle: Der wörtliche Feldname muss in dieser Liste stehen, sonst liest ihn der Generator nie. Anschließend berührte eine erneute Synchronisierung nach einem Monat Änderungen acht von 186 Dateien statt aller, und die Sitemap sagt etwas Wahres.

Dieses Datum erfüllt inzwischen Aufgaben über die Sitemap hinaus. Wie die Seite zur Auffindbarkeit durch KI-Agenten beschreibt, meldet jedes Deployment seine Änderungen über IndexNow, ein Protokoll, an dem Bing und mehrere andere Suchmaschinen teilnehmen. Das Deployment-Skript vergleicht zur Änderungserkennung das `lastmod` jeder Seite mit dem zuletzt gesendeten Wert. Ein Deployment mit einer geänderten Seite meldet deshalb eine Seite. Google steht nicht auf IndexNows eigener Liste teilnehmender Suchmaschinen und erfährt von der Änderung beim nächsten Lesen der Sitemap. Eine Build-Prüfung hält die Datumsangaben korrekt: Sie vergleicht jedes Sitemap-Datum mit dem Verlauf seiner Seite und verweigert die Veröffentlichung, wenn eines falsch ist. Für Menschen statt für Programme gibt es außerdem einen lesbaren Index aller Seiten.

Wo es schiefgeht

Jede Seite behauptet, heute geändert worden zu sein. Die Sitemap verwendet für Hunderte Seiten ein Datum oder eines pro Deployment. Eine Suchmaschine lernt daraus nichts über neue Arbeiten. Die Ursache ist ein Datum aus dem Dateisystem oder der Build-Uhr. Google verwendet `<lastmod>` nur bei durchgängiger, überprüfbarer Genauigkeit, die ein gemeinsamer Zeitstempel sichtbar verfehlt. Nehmen Sie das Datum aus dem Verlauf des Inhalts und vergleichen Sie nach jeder Änderung am Build eine Stichprobe von Einträgen damit.

Die Liste fordert unerwünschte Seiten an. Weiterleitungen, doppelte Adressen mit Tracking-Parametern, mit `noindex` markierte Seiten, Staging-Hosts und Kopien mit einem Canonical-Verweis auf eine andere Seite landen in Sitemaps eines Generators, der jede erzeugte Datei auflistet. Jeder Eintrag schickt einen Crawler zu einer Adresse, von der sich die Seite anschließend selbst lossagt.

Listen Sie nur die kanonische Adresse jeder für Suchergebnisse bestimmten Seite auf. Lassen Sie Sitemap und Canonical-Tag derselben Regel folgen, wie es die Vorlagen dieser Website tun.

Einträge liegen außerhalb des erlaubten Bereichs. Eine Sitemap unter `/blog/sitemap.xml` listet Shopseiten auf oder schreibt relative Pfade wie `/page.html` . Das Protokoll beschränkt eine Datei auf Adressen in oder unter ihrem eigenen Ordner, und Google crawlt Adressen genau wie angegeben. Veröffentlichen Sie die Sitemap an der Wurzel, schreiben Sie jedes `<loc>` als vollständige `https://` -Adresse desselben Hosts und wechseln Sie zum Sitemap-Index, bevor die Liste in die Nähe von 50.000 Einträgen oder 50 MB kommt.

Seit Monaten gelistete Seiten fehlen im Index. Der Eigentümer versteht die Sitemap als Anweisung und wartet. Google nennt sie einen Hinweis: Eine gelistete Adresse kann aus Gründen außerhalb der Datei trotzdem übersprungen oder mit einem Duplikat zusammengeführt werden. Prüfen Sie die Seite statt der Liste: Verlinken andere Seiten sie? Verweist ihr Canonical auf sie selbst? Halten robots.txt oder eine `noindex` -Regel einen Crawler fern?

Weiterführende Lektüre

- Sitemaps-XML-Format — Das Protokoll selbst: alle Tags, Größenlimits, Indexdateien und die Adressen, die eine Datei auflisten darf.
- Eine Sitemap erstellen und einreichen — Googles Regeln, einschließlich der ignorierten Tags und der Voraussetzungen für ein vertrauenswürdiges Datum.
- Informationen zu Sitemaps — Wann Google eine Sitemap für nötig hält und wann gute Verlinkung genügt.
- IndexNow-Dokumentation — Das Protokoll, das teilnehmenden Suchmaschinen Änderungen am Tag ihrer Veröffentlichung meldet.
- RFC 9309: Robots Exclusion Protocol — Der robots.txt-Standard, der einem Crawler erlaubt, neben dessen Regeln eine Zeile `Sitemap:` zu lesen.

Glossar

Für Ihre tatsächlichen Rahmenbedingungen auswählen

Beschreiben Sie den Arbeitsablauf, die vorhandenen Schnittstellen und die Nutzer, die die Veränderung benötigen. Wir können eine prüfbare Lieferung definieren und die Übergabe so gestalten, dass Ihr Team die gelieferten Bestandteile warten kann.

Preise ansehen Kontakt aufnehmen →

Die Begriffe praktisch nutzbar machen

Fragen Sie, wie „Was ist eine Sitemap?“ auf Ihre Organisation zutrifft. Ein konkretes Beispiel aus Ihrem bestehenden System reicht aus, um das Gespräch zu beginnen.

Unsere Leistungen entdecken

welcome@engineer.company +45 50 10 61 84 Kopenhagen, Dänemark

Nehmen Sie diese Seite mit

Kontaktkarte

Mit Sorgfalt und Liebe gestaltet von D.

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

Copyright © 2025 – heute · Engineer ApS · Alle Rechte vorbehalten.

Gedruckt von https://engineer.company/de/glossary/sitemap/