Engineer

Sitemap

en liste over et sites sider, skrevet til maskiner

Hvad det er

Et sitemap er en fil, et website udgiver om sig selv: de adresser, ejeren vil have søgemaskinerne til at kende, én post pr. side og hver med nogle få valgfrie oplysninger. Filen findes, fordi en crawler ellers lærer et site at kende på den langsomme måde, ved at følge links fra sider, den allerede kender, og nogle sider ligger for enden af en lang kæde af links eller har næsten ingen links ind.

Det format, de store søgemaskiner deler, er Sitemaps-protokollen fra sitemaps.org. XML-udgaven pakker listen ind i et `<urlset>`-element med én `<url>` pr. side. Indeni holder `<loc>` den fulde adresse og er det eneste påkrævede felt. `<lastmod>` er datoen for sidens seneste ændring, `<changefreq>` et gæt på, hvor ofte den vil ændre sig, og `<priority>` en værdi fra 0,0 til 1,0, der rangerer siden mod sitets egne andre sider og aldrig mod andres. Protokollen tager også imod en ren tekstfil med én adresse pr. linje eller et RSS- eller Atom-feed.

Protokollen sætter faste grænser. En fil rummer højst 50.000 adresser og 50 MB ukomprimeret, må pakkes med gzip og skal være UTF-8 med escapede værdier, og hver adresse holder sig under 2.048 tegn. Et større site deler listen op og udgiver et sitemap-indeks, en anden slags fil, der navngiver op til 50.000 sitemaps på samme site. Et sitemap må desuden kun liste adresser på eller under sin egen placering: en fil på `/catalog/sitemap.xml` dækker kun siderne under `/catalog/`.

En søgemaskine finder filen på en af to måder. En `Sitemap:`-linje i robots.txt giver dens fulde adresse uafhængigt af reglerne for de enkelte crawlere, eller ejeren indsender den gennem søgemaskinens egne værktøjer, for eksempel Google Search Console.

Ordet bruges også om en side med links til menneskelige besøgende; når forskellen betyder noget, siger man »XML-sitemap« om den, der er skrevet til programmer.

Brug begrebet i en beslutning

Brug »Hvad er et sitemap?« i en konkret beslutning. Vi kan knytte begrebet til det system, I bygger eller køber. Et site eller en app, som de AI-assistenter, jeres kunder bruger, kan finde, læse og handle på, kontrolleret ved hver udgivelse i stedet for lovet én gang.

Se vores ydelser Kontakt os →

Undersøg antagelserne bag ordet

Brug forklaringen og det tilknyttede arbejde til at formulere spørgsmål til jeres projekt. Vi kan undersøge antagelserne sammen, før et værktøj vælges. Har bygget et Hugo-site på 3 sprog — 108 skabelonfiler, der betjener tre sprog og fire formater

Har bygget et Hugo-site på 3 sprog Kontakt os →

Hvorfor det betyder noget

Googles dokumentation er åben om grænserne. Et site, hvis sider er ordentligt linket, kan som regel findes uden et sitemap, og Google sætter grænsen for »lille« ved omkring 500 sider, der skal kunne vises i resultaterne. Google skriver også, at et sitemap ikke garanterer, at alt i det bliver crawlet og indekseret, og kalder en indsendelse et hint og intet mere. Mest gavn gør det på store sites, nye sites med få links udefra og sites med meget video, mange billeder eller nyheder.

Næsten alle sitegeneratorer skriver et sitemap gratis, så den egentlige beslutning er, hvad filen påstår, og hver post er en påstand, som en søgemaskine kan holde op mod selve siden.

Hvilke adresser hører til? Google beder om hver sides kanoniske adresse. En liste fyldt op med omdirigeringer, dubletter eller sider markeret `noindex` beder om netop det, resten af sitet siger, at man skal lade være.

Hvilke felter? Google siger, at den ignorerer `<priority>` og `<changefreq>`. Den bruger kun `<lastmod>`, hvis datoen er konsekvent og efterprøveligt korrekt, for eksempel når den sammenlignes med sidens seneste ændring, og protokollen siger, at datoen skal være den dag, siden sidst blev ændret, og ikke den dag, sitemappet blev genereret. Google regner en ændring af hovedindholdet, de strukturerede data eller linkene som væsentlig og et nyt årstal i copyright-linjen som ubetydeligt.

Det er her, omkostningen ligger. En dato fra filsystemet er gratis og som regel forkert, fordi udrulninger, eksporter og friske checkouts rører filer, hvis tekst er den samme. En dato fra versionshistorikken er ærlig, men så skal buildet have historikken ved hånden og læse den i den rigtige rækkefølge. En dato skrevet i hånden er præcis, indtil nogen glemmer at rette den.

Sådan ser det ud i praksis

Dette sites robots.txt nævner én adresse, `https://engineer.company/sitemap.xml`, og den fil er et sitemap-indeks, der navngiver sitemap-filerne for de offentliggjorte sprogudgaver. Hver post for en side har adressen og en dato samt links til den samme side i dens andre offentliggjorte oversættelser. `<priority>` og `<changefreq>` er udeladt, fordi Google siger, at den ignorerer begge, og protokollen kræver ingen af dem. Da denne artikel blev skrevet, så posten for siden om AI-parathed sådan ud, uden sproglinkene:

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

Listen er selektiv med vilje. Virksomhedens fokuserede satellitsites genudgiver arbejde herfra med en canonical, der peger tilbage, og den samme skabelon holder de kopier ude af hvert satellitsites sitemap: en post beder om at få netop den adresse indekseret, og at bede om en kopi og samtidig pege på en anden side som originalen ville modsige sig selv. Hver vært melder kun sine egne sider.

Datoerne har deres egen historie. På et tidspunkt fortalte sitemappet søgemaskiner, at 172 af dets 176 sider var ændret i samme øjeblik. Årsagen var hverdagsagtig: datoen kom fra hver fils ændringstidspunkt, og indholdseksporten omskriver hver fil ved hver kørsel. Rettelsen flyttede kilden til versionshistorikken med en prioriteret rækkefølge, der først prøver et eksplicit felt, så commit-historikken, så filen. Én fælde er værd at kende: det bogstavelige feltnavn skal stå på listen, ellers læser generatoren det aldrig. Bagefter rørte en gensynkronisering efter en måneds redigeringer otte filer af 186 i stedet for dem alle, og sitemappet siger noget sandt.

Den dato gør nu mere end at stå i sitemappet. Som siden om at være klar til AI-agenter beskriver, melder hver udrulning ændringerne via IndexNow, en protokol, som Bing og flere andre søgemaskiner er med i. Udrulningsscriptet afgør, hvad der er ændret, ved at sammenligne hver sides `lastmod` med det, det sidst sendte, så en udrulning, der ændrer én side, melder én side. Google står ikke på IndexNows egen liste over deltagende søgemaskiner og hører derfor om ændringen, næste gang den læser sitemappet. Den samme side beskriver værnet: kvalitetsporten holder hver sitemap-dato op mod sidens historik og nægter at udgive, når en af dem er forkert. Til mennesker frem for programmer findes også et læsbart indeks over hver side.

Hvor det går galt

Hver side påstår at være ændret i dag. Sitemappet har én dato, eller én pr. udrulning, på tværs af hundredvis af sider, og en søgemaskine, der læser det, lærer intet om, hvilket arbejde der er nyt. Årsagen er en dato fra filsystemet eller buildets ur. Google bruger kun `<lastmod>`, når datoen er konsekvent og efterprøveligt korrekt, og et fælles tidsstempel er synligt ikke det. Hent datoen fra indholdets egen historik, og hold et udsnit af posterne op mod den efter hver ændring af buildet.

Listen beder om sider, sitet vil holde ude. Omdirigeringer, dubletter med sporingsparametre, sider markeret `noindex`, staging-værter og kopier, hvis kanoniske adresse peger et andet sted hen, havner i sitemaps fra en generator, der lister hver fil, den har bygget. Hver af dem sender en crawler til en adresse, som siden selv derefter fornægter. List kun den kanoniske adresse for hver side, der skal kunne vises i resultaterne, og lad sitemappet og canonical-tagget svare ud fra én regel, sådan som dette sites skabeloner gør.

Poster ligger uden for det, filen må dække. Et sitemap på `/blog/sitemap.xml` lister butikssider eller skriver relative stier som `/page.html`. Protokollen begrænser en fil til adresser på eller under dens egen mappe, og Google crawler adresser præcis, som de står. Udgiv sitemappet i roden, skriv hver `<loc>` som en fuld `https://`-adresse på samme vært, og skift til et sitemap-indeks, før listen nærmer sig 50.000 poster eller 50 MB.

Sider, der har stået på listen i månedsvis, kommer ikke i indekset. Ejeren læser sitemappet som en ordre og venter. Google kalder det et hint: en listet adresse kan stadig blive sprunget over eller lagt sammen med en dublet af grunde, der

ligger uden for filen. Se på siden frem for listen: om andre sider linker til den, om dens canonical peger på den selv, og om robots.txt eller en `noindex` -regel holder crawlere væk.

Læs mere

- Sitemaps XML format — selve protokollen: hvert tag, størrelsesgrænserne, indeksfiler og hvilke adresser en fil må liste.
- Build and submit a sitemap — Googles regler, herunder de tags den ignorerer, og hvornår den stoler på en dato.
- Learn about sitemaps — hvornår Google mener, et site har brug for et, og hvornår god linkstruktur er nok.
- IndexNow documentation — protokollen, der fortæller de deltagende søgemaskiner om en ændring samme dag, den går i luften.
- RFC 9309: Robots Exclusion Protocol — standarden for robots.txt, som lader en crawler læse en `Sitemap:` -linje ved siden af reglerne.

Ordliste

Vælg ud fra jeres faktiske vilkår

Beskriv arbejdsgangen, de eksisterende grænseflader og de brugere, der har brug for ændringen. Vi kan definere en testbar leverance og en overdragelse, der gør teamet i stand til at vedligeholde løsningen.

Se, hvad det koster Kontakt os →

Gør begreberne praktiske

Spørg, hvordan »Hvad er et sitemap?« gælder i jeres organisation. Et konkret eksempel fra jeres nuværende system er nok til at begynde samtalen.

Se vores ydelser

welcome@engineer.company +45 50 10 61 84 København, Danmark

Tag denne side med dig

Visitkort

Lavet med kærlighed af D.

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

Copyright © 2025 – nu · Engineer ApS · Alle rettigheder forbeholdes.

Udskrevet fra https://engineer.company/da/glossary/sitemap/