Engineer

Sitemap

en liste over et nettsteds sider, skrevet for maskiner

Hva det er

Et sitemap er en fil et nettsted publiserer om seg selv, med adressene eieren vil at søkemotorene skal kjenne til: én oppføring per side, og et par valgfrie opplysninger knyttet til hver. Ordet er lånt rett fra engelsk og brukes i begge kjønn, sitemapen og sitemapet. Filen finnes fordi en søkerobot ellers lærer et nettsted å kjenne den langsomme veien, ved å følge lenker fra sider den allerede vet om, og noen sider ligger ytterst i en lang kjede av lenker eller har knapt noen lenker inn i det hele tatt.

Formatet de store søkemotorene deler, er Sitemaps-protokollen, publisert på sitemaps.org. I XML-versjonen pakkes listen inn i et `<urlset>` -element med ett `<url>` per side. Inni det holder `<loc>` den fullstendige adressen, og det er det eneste underelementet som er påkrevd. `<lastmod>` er datoen siden sist ble endret, `<changefreq>` en gjetning på hvor ofte den kommer til å endres, og `<priority>` en verdi fra 0,0 til 1,0 som rangerer siden mot nettstedets egne andre sider og aldri mot noen andres. Protokollen godtar også en ren tekstfil med én adresse per linje, eller en RSS- eller Atom-feed.

Protokollen setter faste grenser. Én fil holder høyst 50 000 adresser og 50 MB ukomprimert, kan komprimeres med gzip og må være UTF-8 med spesialtegn kodet som XML-entiteter, og hver adresse holder seg under 2 048 tegn. Et større nettsted deler listen opp og publiserer en sitemapindeks, en egen filtype som navngir opptil 50 000 sitemap-filer på det samme nettstedet. Et sitemap kan dessuten bare liste adresser på eller under sin egen plassering: en fil på `/catalog/sitemap.xml` dekker sidene under `/catalog/` og stopper ved kanten av den mappen.

En søkemotor får vite hvor filen ligger, på én av to måter. En `Sitemap:` -linje i robots.txt oppgir den fullstendige adressen, uavhengig av gruppene for hver user-agent, eller eieren sender den inn gjennom søkemotorens egne verktøy, f.eks. Google Search Console.

Det samme ordet brukes også om en side med lenker for menneskelige besøkende. På norsk heter den gjerne nettstedskart, og når forskjellen betyr noe, sier folk «XML-sitemap» om filen som er skrevet for programmer.

Bruk begrepet i en beslutning

Bruk «Hva er et sitemap?» i en konkret beslutning. Vi kan knytte begrepet til systemet dere bygger eller kjøper. Et nettsted eller en app som AI-assistentene kundene dine bruker, kan finne, lese og handle ut fra, kontrollert ved hver utgivelse i stedet for lovet én gang.

Se tjenestene våre Ta kontakt →

Undersøk antakelsene bak ordet

Bruk forklaringen og det tilknyttede arbeidet til å formulere spørsmål for prosjektet. Vi kan undersøke antakelsene sammen før et verktøy velges. Bygde et flerspråklig Hugo-nettsted på 3 språk — 108 malfiler som betjener tre språk og fire formater

Bygde et flerspråklig Hugo-nettsted på 3 språk Ta kontakt →

Hvorfor det er viktig

Googles dokumentasjon er åpen om begrensningene. Et nettsted der sidene lenker ordentlig til hverandre, blir som regel funnet uten et sitemap, og Google setter grensen for «lite» ved omtrent 500 sider som skal vises i resultatene. Dokumentasjonen sier også at et sitemap ikke garanterer at alt i det blir gjennomsøkt og indeksert, og at det å sende det inn bare er et hint. Det hjelper mest på store nettsteder, på nye nettsteder med få lenker fra andre steder, og på nettsteder med mye video, bilder eller nyheter.

Nesten hver nettstedsgenerator skriver et sitemap gratis, så den egentlige beslutningen er hva filen påstår, og hver oppføring er en påstand en søkemotor kan sjekke mot selve siden.

Hvilke adresser hører hjemme der? Google ber om den kanoniske URL-en til hver side. En liste fylt opp med videresendinger, duplikater eller sider merket `noindex` ber om nøyaktig det resten av nettstedet sier skal stå i fred.

Hvilke felt? Google sier at det overser `<priority>` og `<changefreq>`. `<lastmod>` bruker Google bare når datoen er konsekvent og etterprøvbart riktig, f.eks. når den stemmer med siste endring på selve siden, og protokollen sier at datoen skal

være den dagen siden det lenkes til, sist ble endret, og ikke den dagen sitemapet ble generert. Google regner en endring i hovedinnholdet, de strukturerte dataene eller lenkene som vesentlig, og et nytt årstall i linjen om opphavsrett som ubetydelig.

Det er der kostnaden ligger. En dato fra filsystemet er gratis og som regel feil, fordi produksjonssettinger, eksporter og ferske utsjekkinger rører filer der ordene er de samme. En dato fra versjonshistorikken er ærlig, men da må bygget ha den historikken tilgjengelig og lese den i riktig rekkefølge. En dato skrevet inn for hånd er presis helt til noen glemmer å endre den.

Slik ser det ut i praksis

Filen robots.txt på dette nettstedet navngir én adresse, `https://engineer.company/sitemap.xml`, og den filen er en sitemapindeks som navngir sitemap-filene for de publiserte språkutgavene. Hver sideoppføring har adressen og en dato, pluss lenker til den samme siden i de andre publiserte oversettelsene. Feltene `<priority>` og `<changefreq>` er utelatt, siden Google sier at det overser begge, og protokollen krever ingen av dem. Da denne artikkelen ble skrevet, så oppføringen for siden om AI-beredskap slik ut, med språklenkene lagt til side:

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

Listen er selektiv med vilje. Selskapets fokuserte satellittnettsteder publiserer arbeid herfra på nytt med en kanonisk URL som peker tilbake, og den samme malen holder de kopiene utenfor sitemapet til hver satellitt: en oppføring ber en søkemotor indeksere nøyaktig den adressen, og å be om en kopi og samtidig navngi en annen side som originalen ville motsi seg selv. Hver vert melder bare sine egne sider.

Datoene har sin egen historie. På et tidspunkt fortalte sitemapet søkemotorene at 172 av de 176 sidene var endret i samme øyeblikk. Årsaken var hverdagslig: datoen kom fra endringstidspunktet til hver fil, og innholdseksporten skriver om hver fil ved hver kjøring. Rettelsen flyttet kilden til versjonshistorikken, med en reservekjede som først prøver et eksplisitt felt, så commit-historikken, så filen.

Den hadde én felle det er verdt å kjenne til: det bokstavelige feltnavnet må stå i den listen, ellers leser generatoren det aldri. Etterpå berørte en ny synkronisering etter en måned med endringer åtte filer av 186 i stedet for alle, og sitemapet sier noe sant.

Den datoen gjør nå nytte også utenfor sitemapet. Som siden om å være klar for AI-agenter beskriver, melder hver produksjonssetting fra om hva som er endret via IndexNow, en protokoll som Bing og flere andre søkemotorer er med i. Skriptet for produksjonssetting avgjør hva som er endret ved å sammenligne `lastmod` for hver side med det det sendte sist, så en produksjonssetting som endrer én side, melder én side. Google står utenfor IndexNows egen liste over deltakende søkemotorer, så Google får vite om endringen neste gang det leser sitemapet. Den samme siden nevner vernet: bygget holder hver dato i sitemapet opp mot sidens historikk og nekter å publisere når en av dem er feil. For mennesker og ikke programmer finnes det også et lesbart nettstedskart over hver side.

Hvor det går galt

Hver side påstår å være endret i dag. Sitemapet har én dato, eller én per produksjonssetting, over hundrevis av sider, og en søkemotor som leser det, lærer ingenting om hvilket arbeid som er nytt. Årsaken er en dato hentet fra filsystemet eller fra klokken under bygget. Google bruker `<lastmod>` bare når datoen er konsekvent og etterprøvbart riktig, og et felles tidsstempel er synlig noe annet. Hent datoen fra innholdets egen historikk, og sammenlign et utvalg av oppføringene med den etter hver endring i bygget.

En liste som ber om sider nettstedet vil ha holdt utenfor. Videresendinger, dupliserte adresser med sporingsparametere, sider merket `noindex`, testverter og kopier der den kanoniske URL-en peker et annet sted, havner alle i sitemaps skrevet av en generator som lister hver fil den bygde. Hver av dem sender en søkerobot til en adresse som siden selv deretter tar avstand fra. List bare den kanoniske URL-en til hver side som er ment for resultatene, og la sitemapet og den kanoniske taggen svare ut fra én regel, slik malene på dette nettstedet gjør.

Oppføringer utenfor det filen har lov til å dekke. Et sitemap på `/blog/sitemap.xml` lister sider fra nettbutikken, eller skriver relative stier som `/page.html`. Protokollen begrenser en fil til adresser på eller under dens egen mappe, og Google gjennomsøker adressene nøyaktig slik de står. Publiser

sitemapet i roten, skriv hver `<loc>` som en fullstendig `https://`-adresse på samme vert, og gå over til en sitemapindeks før listen nærmer seg 50 000 oppføringer eller 50 MB.

Sider som har stått på listen i månedsvis og fortsatt mangler i indeksen. Eieren leser sitemapet som en instruks og venter. Google kaller det et hint: en adresse på listen kan likevel bli hoppet over eller slått sammen med et duplikat av grunner som ligger utenfor filen. Se på siden i stedet for listen: om andre sider lenker til den, om den kanoniske URL-en peker på siden selv, og om robots.txt eller en `noindex`-regel holder søkeroboten unna.

Les mer

- Sitemaps XML format — selve protokollen: hver tagg, størrelsesgrensene, indeksfilene og hvilke adresser en fil har lov til å liste.
- Build and submit a sitemap — Googles regler, med taggene Google overser og når en dato blir trodd.
- Learn about sitemaps — når Google mener et nettsted trenger et sitemap, og når god lenking holder.
- IndexNow documentation — protokollen som forteller deltakende søkemotorer om en endring samme dag som den går ut.
- RFC 9309: Robots Exclusion Protocol — standarden for robots.txt, som lar en søkerobot lese en `Sitemap:`-linje ved siden av reglene.

Ordliste

Velg ut fra deres faktiske rammer

Beskriv arbeidsflyten, eksisterende grensesnitt og brukerne som trenger endringen. Vi kan definere en testbar leveranse og en overlevering som gjør teamet i stand til å vedlikeholde løsningen.

Se hva det koster Ta kontakt →

Gjør begrepene praktiske

Spør hvordan «Hva er et sitemap?» gjelder i organisasjonen deres. Et konkret eksempel fra dagens system er nok til å starte samtalen.

Se tjenestene våre

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

Ta med deg denne siden

Visittkort

Laget med kjærlighet av D.

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

Opphavsrett © 2025 – nå · Engineer ApS · Alle rettigheter forbeholdt.

Skrevet ut fra https://engineer.company/nb/glossary/sitemap/