Engineer

Webbplatskarta

en lista över en webbplats sidor, skriven för maskiner

Vad det är

En webbplatskarta är en fil som en webbplats publicerar om sig själv: de adresser som ägaren vill att sökmotorer ska känna till, en post per sida, var och en med några valfria uppgifter. I bestämd form blir ordet webbplatskartan, i plural webbplatskartor, och på engelska heter den sitemap. Den finns eftersom en sökrobot annars lär sig en webbplats den långsamma vägen, genom att följa länkar från sidor som den redan känner till, och en del sidor ligger längst ut i en lång kedja av länkar eller har nästan inga inkommande länkar alls.

Formatet som de stora sökmotorerna delar är Sitemaps-protokollet, publicerat på sitemaps.org. Dess XML-version omsluter listan i ett `<urlset>` -element med ett `<url>` per sida. Inuti det håller `<loc>` den fullständiga adressen och är det enda obligatoriska underelementet. `<lastmod>` är datumet då sidan senast ändrades, `<changefreq>` en gissning om hur ofta den kommer att ändras och `<priority>` ett värde från 0,0 till 1,0 som rangordnar sidan mot webbplatsens egna andra sidor, aldrig mot någon annans. Protokollet godtar också en vanlig textfil med en adress per rad, eller ett RSS- eller Atom-flöde.

Protokollet sätter hårda gränser. En fil rymmer högst 50 000 adresser och 50 MB okomprimerat, får komprimeras med gzip och måste vara kodad i UTF-8, med specialtecken i värdena skrivna som XML-entiteter; varje adress ska vara kortare än 2 048 tecken. En större webbplats delar upp sin lista och publicerar ett index över webbplatskartor, på engelska sitemap index, en andra sorts fil som namnger upp till 50 000 webbplatskartor på samma webbplats. En webbplatskarta får dessutom bara lista adresser på eller under sin egen plats: en fil på `/catalog/sitemap.xml` täcker sidorna under `/catalog/` och slutar vid den mappens gräns.

En sökmotor får veta var filen ligger på ett av två sätt. En rad `Sitemap:` i robots.txt anger dess fullständiga adress, oberoende av alla user-agent-grupper, eller så skickar ägaren in den via sökmotorns egna verktyg, t.ex. Google Search Console.

Samma ord betecknar också en sida med länkar för mänskliga besökare. När skillnaden spelar roll säger man XML-webbplatskarta, på engelska ”XML sitemap”, om den som är skriven för program.

Använd begreppet i ett beslut

Använd ”Vad är en webbplatskarta?” i ett konkret beslut. Vi kan knyta begreppet till systemet ni bygger eller köper. En webbplats eller app som de AI-assistenter dina kunder använder kan hitta, läsa och agera på, kontrollerad vid varje release i stället för utlovad en gång.

Se våra tjänster Hör av er →

Granska antagandena bakom ordet

Använd förklaringen och det länkade arbetet för att formulera frågor till projektet. Vi kan undersöka antagandena tillsammans innan ett verktyg väljs. Byggde en flerspråkig Hugo-webbplats på 3 språk — 108 mallfiler som täcker tre språk och fyra format

Byggde en flerspråkig Hugo-webbplats på 3 språk Hör av er →

Varför det spelar roll

Googles dokumentation är öppen med begränsningarna. En webbplats vars sidor är ordentligt länkade kan oftast upptäckas utan en webbplatskarta, och Google drar gränsen för ”liten” vid ungefär 500 sidor som ska synas i resultaten. Där står också att en webbplatskarta inte garanterar att allt som står i den genomsöks och indexeras, och att den som skickas in bara är en ledtråd. Mest nytta gör den för stora webbplatser, för nya webbplatser med få länkar från andra håll och för webbplatser med mycket video, bilder eller nyheter.

Nästan varje webbplatsgenerator skriver en utan extra arbete, så det verkliga beslutet gäller vad filen påstår, och varje post är ett påstående som en sökmotor kan pröva mot själva sidan.

Vilka adresser ska med? Google ber om den kanoniska adressen till varje sida. En lista som fylls ut med omdirigeringar, dubbletter eller sidor märkta `noindex` ber om just det som resten av webbplatsen säger åt sökmotorn att låta bli.

Vilka fält? Enligt Google ignoreras `<priority>` och `<changefreq>`. `<lastmod>` används bara om datumet genomgående och kontrollerbart stämmer, t.ex. jämfört med när sidan senast ändrades, och protokollet säger att datumet ska

vara det då den länkade sidan senast ändrades, inte det då webbplatskartan genererades. Google räknar en ändring av huvudinnehållet, den strukturerade datan eller länkarna som betydande, och ett nytt årtal i upphovsrättsraden som obetydligt.

Det är där kostnaden ligger. Ett datum från filsystemet är gratis och oftast fel, eftersom driftsättningar, exporter och nya utcheckningar rör filer vars text förblev densamma. Ett datum ur versionshistoriken är ärligt, men då behöver bygget ha den historiken på plats och läsa den i rätt ordning. Ett datum som skrivs in för hand är exakt tills någon glömmer att ändra det.

Så ser det ut i praktiken

Den här webbplatsens robots.txt namnger en enda adress, `https://engineer.company/sitemap.xml`, och den filen är ett index som namnger webbplatskartorna för de publicerade språkversionerna. Varje post för en sida innehåller adressen och ett datum, plus länkar till samma sida i dess övriga publicerade översättningar. `<priority>` och `<changefreq>` är utelämnade, eftersom Google säger sig ignorera båda och protokollet inte kräver något av dem. När den här artikeln skrevs såg posten för sidan om att vara redo för AI-agenter ut så här, med språklänkarna borträknade:

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

Listan är selektiv med avsikt. Företagets fokuserade satellitwebbplatser publicerar arbete härifrån på nytt med en kanonisk URL som pekar tillbaka hit, och samma mall håller de kopiorna utanför varje satellitwebbplats egen webbplatskarta: en post ber en sökmotor att indexera exakt den adressen, och att be om en kopia samtidigt som man pekar ut en annan sida som originalet vore att säga emot sig själv. Varje webbplats meddelar bara sina egna sidor.

Datumen har en egen historia. Vid ett tillfälle talade webbplatskartan om för sökmotorer att 172 av dess 176 sidor hade ändrats i samma ögonblick. Orsaken var vardaglig: datumet kom från varje fils ändringstid, och innehållsexporten skriver om varje fil vid varje körning. Åtgärden flyttade källan till versionshistoriken, med en reservkedja som först prövar ett explicit fält, sedan

commit-historiken och sist filen. Den innehöll en fälla som är värd att känna till: det bokstavliga fältnamnet måste stå i listan, annars läser generatorn aldrig fältet. Efteråt berörde en ny synkronisering efter en månads ändringar åtta filer av 186 i stället för samtliga, och webbplatskartan säger något som är sant.

Det datumet gör nu nytta även utanför webbplatskartan. Som sidan om att vara redo för AI-agenter beskriver meddelar varje driftsättning vad som har ändrats via IndexNow, ett protokoll som Bing och flera andra sökmotorer deltar i. Skriptet för driftsättning avgör vad som har ändrats genom att jämföra varje sidas `lastmod` med det som det senast skickade, så en driftsättning som ändrar en sida meddelar en sida. Google står inte på IndexNows egen lista över deltagande sökmotorer och får därför veta om ändringen nästa gång webbplatskartan läses. Samma sida beskriver skyddet: bygget läser varje datum i webbplatskartan mot sidans historik och vägrar publicera när något av dem är fel. För människor snarare än program finns också en läsbar förteckning över varje sida.

Var det går fel

Varje sida påstår att den ändrades i dag. Webbplatskartan bär ett enda datum, eller ett per driftsättning, för hundratals sidor, och en sökmotor som läser den lär sig ingenting om vilket arbete som är nytt. Orsaken är ett datum som har hämtats från filsystemet eller från byggets klocka. Google använder `<lastmod>` bara när datumet genomgående och kontrollerbart stämmer, vilket en gemensam tidsstämpel uppenbart inte gör. Hämta datumet ur innehållets egen historik, och jämför ett urval av posterna mot den efter varje ändring av bygget.

Listan ber om sidor som webbplatsen vill hålla utanför. Omdirigeringar, dubbla adresser med spårningsparametrar, sidor märkta `noindex`, staging-miljöer och kopior vars kanoniska adress pekar någon annanstans hamnar alla i webbplatskartor som skrivs av en generator som listar varje fil den har byggt. Var och en skickar en sökrobot till en adress som sidan själv sedan tar avstånd från. Lista bara den kanoniska adressen till varje sida som är avsedd för resultaten, och låt webbplatskartan och den kanoniska taggen svara utifrån en och samma regel, så som den här webbplatsens mallar gör.

Poster hamnar utanför det som filen får täcka. En webbplatskarta på `/blog/sitemap.xml` listar sidor i en webbutik, eller skriver relativa sökvägar som `/page.html`. Protokollet begränsar en fil till adresser på eller under dess egen mapp, och Google genomsöker adresser exakt som de är skrivna. Publicera

webbplatskartan i roten, skriv varje `<loc>` som en fullständig `https://`-adress på samma värdnamn, och gå över till ett index över webbplatskartor innan listan närmar sig 50 000 poster eller 50 MB.

Sidor som har listats i månader stannar utanför indexet. Ägaren läser webbplatskartan som en instruktion och väntar. Google kallar den en ledtråd: en listad adress kan ändå hoppas över eller slås ihop med en dubblett av skäl som ligger utanför filen. Titta på sidan i stället för på listan: om andra sidor länkar till den, om dess kanoniska adress pekar på den själv och om robots.txt eller en `noindex`-regel håller sökrobotar borta.

Läs mer

- Sitemaps XML format — själva protokollet: varje tagg, storleksgränserna, indexfilerna och vilka adresser en fil får lista.
- Build and submit a sitemap — reglerna från Google, bl.a. vilka taggar som ignoreras och när ett datum litas på.
- Learn about sitemaps — när Google säger att en webbplats behöver en och när god länkning räcker.
- IndexNow documentation — protokollet som berättar för deltagande sökmotorer om en ändring samma dag som den levereras.
- RFC 9309: Robots Exclusion Protocol — standarden för robots.txt, som låter en sökrobot läsa en rad `Sitemap:` bredvid sina regler.

Ordlista

Välj utifrån era faktiska förutsättningar

Beskriv arbetsflödet, befintliga gränssnitt och användarna som behöver förändringen. Vi kan definiera en testbar leverans och en överlämning som gör att teamet kan underhålla lösningen.

Se vad det kostar Hör av er →

Gör begreppen praktiska

Fråga hur ”Vad är en webbplatskarta?” gäller i er organisation. Ett konkret exempel från ert nuvarande system räcker för att börja samtalet.

Se våra tjänster

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

Ta med dig den här sidan

Visitkort

Omsorgsfullt skapad med kärlek av D.

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

Upphovsrätt © 2025 – nu · Engineer ApS · Alla rättigheter förbehållna.

Utskrivet från https://engineer.company/sv/glossary/sitemap/