Engineer

mapa del sitio

Lista de páginas de un sitio destinada a las máquinas

Qué es

Un mapa del sitio es un archivo que un sitio web publica sobre sí mismo: las direcciones que su propietario quiere dar a conocer a los buscadores, con una entrada por página y algunos datos opcionales en cada una. Existe porque, de lo contrario, un rastreador descubre un sitio por la vía lenta, siguiendo enlaces desde páginas que ya conoce, y algunas páginas están al final de una larga cadena de enlaces o apenas reciben enlaces entrantes.

El formato que comparten los grandes buscadores es el protocolo Sitemaps, publicado en sitemaps.org. Su versión XML envuelve la lista en un elemento `<urlset>` con un `<url>` por página. Dentro, `<loc>` contiene la dirección completa y es el único elemento hijo obligatorio. `<lastmod>` es la fecha del último cambio de la página, `<changefreq>` una estimación de la frecuencia con la que cambiará, y `<priority>` un valor de 0,0 a 1,0 que ordena la página frente a las demás páginas del propio sitio, nunca frente a las de otros. El protocolo también acepta un archivo de texto sin formato con una dirección por línea, o un canal RSS o Atom.

El protocolo establece límites estrictos. Un archivo contiene como máximo 50.000 direcciones y 50 MB sin comprimir, puede comprimirse con gzip y debe estar en UTF-8 con los caracteres de sus valores escapados mediante entidades; cada dirección debe tener menos de 2.048 caracteres. Un sitio más grande divide su lista y publica un índice de mapas del sitio, un segundo tipo de archivo que identifica hasta 50.000 mapas del mismo sitio. Además, un mapa solo puede incluir direcciones en su propia ubicación o por debajo de ella: un archivo en `/catalog/sitemap.xml` abarca las páginas de `/catalog/` y se detiene en el límite de esa carpeta.

Un buscador descubre dónde está el archivo de dos maneras. Una línea `Sitemap:` en robots.txt indica su dirección completa, con independencia de cualquier grupo de agentes de usuario, o el propietario lo envía mediante las herramientas del propio buscador, como Google Search Console.

La misma expresión también designa una página de enlaces para visitantes; cuando importa distinguirlos, se habla de «mapa del sitio XML» para el archivo destinado a programas.

Utilice el concepto en una decisión

Aplique «¿Qué es un mapa del sitio?» a una decisión real. Podemos ayudarle a relacionar el concepto con el sistema que está construyendo o comprando. Un sitio o una aplicación que los asistentes de IA de sus clientes puedan encontrar, leer y utilizar, con comprobaciones en cada versión en lugar de una promesa inicial.

Explore nuestros servicios Póngase en contacto →

Compruebe los supuestos detrás del término

Utilice la explicación y los trabajos enlazados para formular preguntas sobre su proyecto. Podemos examinar los supuestos con usted antes de elegir una herramienta. Sitio Hugo multilingüe en 3 idiomas — 108 archivos de plantillas para tres idiomas y cuatro formatos

Sitio Hugo multilingüe en 3 idiomas Póngase en contacto →

Por qué importa

La documentación de Google es clara sobre los límites. Un sitio cuyas páginas están bien enlazadas normalmente se puede descubrir sin un mapa del sitio, y Google sitúa el límite de un sitio «pequeño» en unas 500 páginas que deban aparecer en los resultados. También dice que un mapa del sitio «no garantiza que se rastreen e indexen todos los elementos de su mapa del sitio» y que enviarlo «es solo una sugerencia». Donde más ayuda es en sitios grandes, sitios nuevos con pocos enlaces desde otros sitios y sitios con mucho vídeo, imágenes o noticias.

Casi todos los generadores de sitios crean uno sin coste, así que la verdadera decisión es qué afirma el archivo, y cada entrada es una declaración que un buscador puede contrastar con la propia página.

¿Qué direcciones se incluyen? Google pide la dirección canónica de cada página. Una lista engrosada con redirecciones, duplicados o páginas marcadas con `noindex` pide exactamente aquello que el resto del sitio indica que se deje de lado.

¿Qué campos? Google dice que ignora `<priority>` y `<changefreq>`. Solo utiliza `<lastmod>` «si es exacto de forma constante y verificable (por ejemplo, comparándolo con la última modificación de la página)», y el protocolo dice que la fecha «debe establecerse en la fecha en que se modificó por última vez la página enlazada, no cuando se genera el mapa del sitio». Google considera significativos los cambios en el contenido principal, los datos estructurados o los enlaces de una página, y trivial un nuevo año de copyright.

Ahí está el coste. Una fecha del sistema de archivos es gratuita y normalmente incorrecta, porque los despliegues, las exportaciones y las nuevas copias del repositorio modifican archivos cuyas palabras no han cambiado. Una fecha del historial de versiones es fiel, pero la compilación necesita disponer de ese historial y leerlo en el orden correcto. Una fecha introducida a mano es precisa hasta que alguien olvida cambiarla.

Cómo se aplica en la práctica

El robots.txt de este sitio indica una dirección, `https://engineer.company/sitemap.xml`, y ese archivo es un índice que enumera los mapas del sitio de las ediciones lingüísticas publicadas. Cada entrada de página incluye la dirección y una fecha, además de enlaces a la misma página en sus otras traducciones publicadas. Se omiten `<priority>` y `<changefreq>`, porque Google afirma que ignora ambos y el protocolo no exige ninguno. Cuando se escribió este artículo, la entrada de la página sobre preparación para la IA tenía este aspecto, dejando aparte sus enlaces a otras lenguas:

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

La selección de la lista es deliberada. Los sitios satélite especializados de la empresa vuelven a publicar trabajos de aquí con una dirección canónica que apunta al original, y la misma plantilla excluye esas copias del mapa de cada satélite: una entrada pide a un buscador que indexe esa dirección exacta, y pedir que indexe una copia mientras se identifica otra página como original sería contradictorio. Cada host anuncia solo sus propias páginas.

Las fechas tienen su propia historia. En un momento dado, el mapa del sitio decía a los buscadores que 172 de sus 176 páginas habían cambiado en el mismo instante. La causa era habitual: la fecha procedía de la hora de modificación de cada archivo, y la exportación de contenido reescribe todos los archivos en cada ejecución. La corrección trasladó el origen al historial de versiones, con una cadena de alternativas que prueba primero un campo explícito, después el historial de commits y, por último, el archivo. Encerraba una trampa que conviene conocer: el nombre literal del campo tiene que figurar en esa lista, o el generador nunca lo lee. Después, una nueva sincronización tras un mes de cambios modificó ocho archivos de 186 en lugar de todos ellos, y el mapa del sitio dice algo cierto.

Esa fecha ahora cumple una función más allá del mapa del sitio. Como explica la página sobre cómo facilitar que los agentes de IA encuentren el sitio, cada despliegue anuncia lo que ha cambiado mediante IndexNow, un protocolo en el que participan Bing y varios buscadores más. El script de despliegue decide qué ha cambiado comparando el `lastmod` de cada página con el que envió por última vez, de modo que un despliegue que modifica una página anuncia una página. Google no figura en la lista de buscadores participantes de IndexNow, por lo que descubre el cambio cuando vuelve a leer el mapa del sitio. Una comprobación de la compilación vela por la exactitud de las fechas: contrasta cada fecha del mapa con el historial de la página e impide publicar si alguna es incorrecta. Para las personas, en lugar de los programas, también hay un índice legible de todas las páginas.

Dónde surgen los problemas

Todas las páginas afirman haber cambiado hoy. El mapa del sitio utiliza una sola fecha, o una por despliegue, para cientos de páginas, y un buscador que lo lea no aprende nada sobre qué trabajo es nuevo. La causa es una fecha tomada del sistema de archivos o del reloj de la compilación. Google solo utiliza `<lastmod>` cuando su exactitud es constante y verificable, algo que una marca de tiempo compartida evidentemente no es. Tome la fecha del historial del propio contenido y contraste una muestra de entradas con él después de cualquier cambio en la compilación.

La lista pide páginas que el sitio quiere excluir. Redirecciones, direcciones duplicadas con parámetros de seguimiento, páginas marcadas con `noindex`, hosts de pruebas y copias cuya dirección canónica apunta a otro lugar aparecen en mapas escritos por un generador que enumera todos los archivos que ha compilado. Cada una envía a un rastreador a una dirección que la propia página

desautoriza después. Incluya solo la dirección canónica de cada página destinada a los resultados y haga que el mapa del sitio y la etiqueta canónica respondan a una misma regla, como hacen las plantillas de este sitio.

Las entradas quedan fuera del ámbito que puede abarcar el archivo. Un mapa en `/blog/sitemap.xml` incluye páginas de la tienda o escribe rutas relativas como `/page.html`. El protocolo limita un archivo a las direcciones situadas en su propia carpeta o por debajo de ella, y Google rastrea las direcciones tal como están escritas. Publique el mapa en la raíz, escriba cada `<loc>` como una dirección `https://` completa en el mismo host y pase a un índice de mapas antes de que la lista se acerque a las 50.000 entradas o los 50 MB.

Las páginas llevan meses en la lista y siguen fuera del índice. El propietario interpreta el mapa del sitio como una instrucción y espera. Google lo llama una sugerencia: una dirección incluida puede omitirse o agruparse con un duplicado por motivos ajenos al archivo. Examine la página en lugar de la lista: si otras páginas la enlazan, si su dirección canónica apunta a sí misma y si robots.txt o una regla `noindex` mantiene alejado a un rastreador.

Lecturas adicionales

- Formato XML de Sitemaps — el propio protocolo: todas las etiquetas, los límites de tamaño, los archivos de índice y las direcciones que puede incluir un archivo.
- Crear y enviar un mapa del sitio — las reglas de Google, incluidas las etiquetas que ignora y cuándo confía en una fecha.
- Información sobre los mapas del sitio — cuándo Google considera necesario un mapa del sitio y cuándo basta con buenos enlaces.
- Documentación de IndexNow — el protocolo que comunica los cambios a los buscadores participantes el mismo día en que se publican.
- RFC 9309: protocolo de exclusión de robots — el estándar de robots.txt, que permite a un rastreador leer una línea `Sitemap:` junto a sus reglas.

Glosario

Elija según sus restricciones reales

Describa el flujo de trabajo, las interfaces existentes y los usuarios que necesitan el cambio. Podemos definir una entrega comprobable y un traspaso que permita a su equipo mantener lo entregado.

Vea cuánto cuesta Póngase en contacto →

Haga práctica la terminología

Pregunte cómo se aplica «¿Qué es un mapa del sitio?» a su organización. Basta un ejemplo concreto de su sistema actual para iniciar la conversación.

Explore nuestros servicios

welcome@engineer.company +45 50 10 61 84 Copenhague, Dinamarca

Llévese esta página

Tarjeta de contacto

Creado con esmero y cariño por D.

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

Derechos de autor © 2025 – actualidad · Engineer ApS · Todos los derechos reservados.

Impreso desde https://engineer.company/es/glossary/sitemap/