



FICHA TÉCNICA • CAPTURAR • FUENTES DE SITIO WEB

¿Todo está en su sitio web? ¿Quién lo vuelve a escribir?

Su sitio web se convierte en el conocimiento de su agente.

Las fuentes de sitio web leen las páginas HTTPS públicas de su sitio web según un calendario. Cada página se convierte en un artículo interno de la base de conocimiento que usted elija, y sus agentes responden a partir de ella por teléfono y en el chat del sitio web.

DE UN VISTAZO

Hasta 2000 páginas por fuente, profundidad de enlaces de hasta 6 y un calendario de cada 6 horas a mensual. Las páginas modificadas se reescriben, y las páginas desaparecidas se retiran tras dos ejecuciones.

01 · DE QUÉ TRATA

Dos versiones de la misma respuesta. Ahora su sitio web fija los datos.

Los horarios, las páginas de producto y los textos de ayuda llevan años en su sitio web. Volver a escribirlos para el agente lleva tiempo, y cada cambio en la web deja la copia desactualizada. Las fuentes de sitio web leen las propias páginas y traen los cambios con la siguiente ejecución.

ENTRADA

URL de inicio con https, sitemap opcional y más URL de inicio en el mismo host

RESULTADO

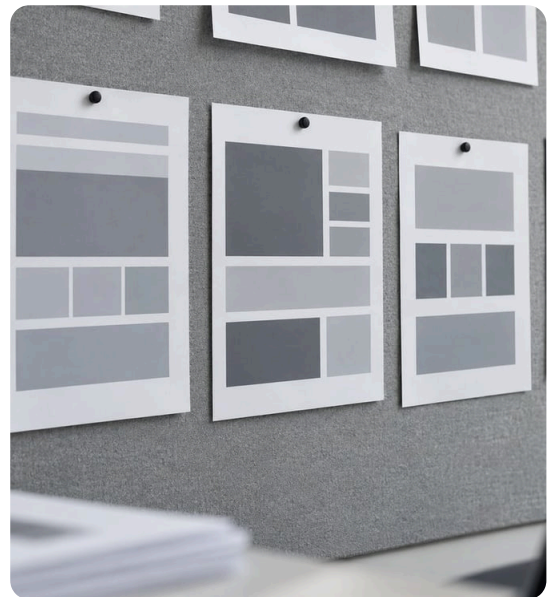
Un artículo interno por página, con la URL de la página como fuente y el idioma detectado

CALENDARIO

Cada 6 o 12 horas, a diario, semanal, cada 14 días o mensual

LECTORES

Agentes de conversación por teléfono y en el chat del sitio web, agentes de chat en Agent AI

**Un host, muchas páginas**

Desde la URL de inicio, el rastreador sigue los enlaces del mismo host, con o sin www. Los sitemaps de robots.txt y un sitemap propio amplían la lista.

Definir el alcance

Páginas máximas y profundidad de enlaces por fuente, además de prefijos de ruta, patrones de exclusión, URL con cadena de consulta y rutas de listado como /tag.

JavaScript incluido

Fast lee el HTML, Full renderiza en un navegador y Auto cambia por sí solo cuando una página parece vacía. Un selector de clic abre, por ejemplo, los acordeones de preguntas frecuentes.

Artículos limpios

La exclusión por CSS mantiene fuera del artículo zonas como la navegación o el pie de página. Si lo desea, el artículo toma un resumen breve de la página.

Las normas del sitio web

Se respetan robots.txt y noindex. Si una página indexada se bloquea más tarde, el rastreador retira su artículo.

Cambios, no reescrituras completas

El rastreador detecta las páginas modificadas por su contenido y deja en paz las que no han cambiado. Las páginas que desaparecen se retiran tras dos ejecuciones sin encontrarlas.

02 · CÓMO FUNCIONA

Añada una fuente, compruébela, asígnela al agente.

Una fuente de sitio web solo necesita unos pocos campos. La vista previa muestra lo que encuentra el rastreador antes de la primera ejecución, y el catálogo muestra después lo que ha ocurrido con cada URL.

1

Añadir una fuente

En **Knowledge Base AI > Fuentes del sitio web**, use **Agregar fuente**: indique el nombre, la URL de inicio, la base de conocimiento de destino y el intervalo de actualización.

2

Comprobar la vista previa

La **Vista previa** muestra robots.txt, sitemaps y candidatos sin indexar nada. Después, define el alcance y el rastreador.

3

Ejecutar y supervisar

La primera ejecución empieza unos minutos después de guardar, y **Sincronizar ahora** programa otra. **Páginas indexadas** muestra el estado y el código HTTP de cada URL.

4

Asignarla al agente

En **Conversation AI > Agentes de conversación**, elija la base de conocimiento en la pestaña **Conocimiento**. A partir de entonces, el agente responde a partir de los artículos del sitio web.

DATOS TÉCNICOS

CARACTERÍSTICA	DETALLES
Entrada	Páginas HTTPS públicas de un solo host, texto HTML
Alcance por fuente	Hasta 2000 páginas (por defecto 200), profundidad de enlaces de hasta 6 (por defecto 3)
Fuentes por cuenta	Hasta 25
Rastreador	Fast, Full o Auto, con selectores de clic y de espera
Salida	Artículo interno por página, con la URL de la página, el idioma detectado y las etiquetas website y host
Resumen	Páginas indexadas con estado, renderizado, HTTP y última lectura; Historial de sincronización por fuente
Acceso y facturación	App web en Knowledge Base AI > Fuentes del sitio web ; facturación según el uso

03 · EN LA PRÁCTICA

Donde el conocimiento del sitio web llega al teléfono y al chat.

Guías para el chat del sitio web

Una empresa de software tiene sus guías bajo /docs. Una fuente de sitio web con el prefijo de ruta /docs las lee a diario, y el chat del sitio web cita la página correspondiente en su respuesta.

Páginas de producto para la línea de atención

Un comercio minorista hace leer sus páginas de producto cada semana. El agente de conversación por teléfono responde preguntas sobre tallas y materiales con los mismos textos que ven los clientes en línea.

Sitio web más sus propios artículos

Todo lo que no está en el sitio web, como un cierre por vacaciones, se añade como artículo propio en la misma base de conocimiento. El agente lee ambos juntos.

PREGUNTAS FRECUENTES

¿Qué páginas lee una fuente de sitio web?

Páginas HTTPS públicas de un solo host, desde la URL de inicio y, si lo desea, mediante sitemaps y más URL de inicio. Lee el texto de las páginas HTML, sin inicio de sesión, cookies ni credenciales.

¿Qué pasa con mis propias ediciones de un artículo del sitio web?

Cuando la página cambia, la siguiente ejecución reescribe el artículo. Por eso, las adiciones deben ir en un artículo propio dentro de la misma base de conocimiento.

¿Qué tan actualizado está el conocimiento del agente?

Tan actualizado como la última ejecución. Usted elige el calendario, de cada 6 horas a mensual, y **Sincronizar ahora** programa una ejecución adicional. Dos ejecuciones manuales están separadas por 15 minutos.

¿Las páginas aparecen en el portal de clientes?

Los artículos del sitio web son internos: su equipo y los agentes de la base de conocimiento los leen, y no aparecen en el portal de clientes.

EMPEZAR

Haga de su sitio web su primera fuente

En la app web, vaya a **Knowledge Base AI > Fuentes del sitio web**, use **Agregar fuente** para indicar la URL de inicio, compruebe la vista previa y guarde. La primera ejecución empieza unos minutos después.

paperoffice.ai/es/ai-web-sources-fuente-conocimiento-agente-rastreador-respuestas



MÁS SOBRE EL MÓDULO

paperoffice.ai/es/ai-web-sources-fuente-conocimiento-agente-rastreador-respuestas

PRECIOS

paperoffice.ai/es/precios

CONTACTO

paperoffice.ai/es/contacto