

PAPEROFFICE
ENTERPRISE./AI

FIȘĂ TEHNICĂ · CAPTURARE · SURSE DIN SITE-UL WEB

Totul e pe site. Cine îl mai retastează? Site-ul dumneavoastră devine cunoștințele agentului.

Sursele de pe site citesc după program paginile publice HTTPS ale site-ului dumneavoastră. Fiecare pagină devine un articol intern în baza de cunoștințe aleasă de dumneavoastră, iar agenții răspund din ea la telefon și în chatul de pe site.

PE SCURT

Până la 2000 de pagini pe sursă, adâncime de link până la 6, un ritm de la fiecare 6 ore până lunar. Paginile modificate se rescriu, cele care dispar se retrag după două rulări.

01 · DESPRE CE ESTE VORBA

Două versiuni ale aceluiași răspuns. Acum site-ul stabilește faptele.

Programul, paginile de produse și textele de ajutor stau de mult pe site-ul dumneavoastră. Să le retastați pentru agent înseamnă timp, iar fiecare modificare de pe site învechește copia. Sursele de pe site citesc singure paginile și aduc modificările odată cu următoarea rulare.

INTRARE
URL de start cu https, opțional sitemap și alte URL-uri de start de pe același host

REZULTAT
Un articol intern pe pagină, cu URL-ul paginii ca sursă și limba detectată

PROGRAM
La fiecare 6 sau 12 ore, zilnic, săptămânal, la 14 zile sau lunar

CITITORI
Agenți conversaționali la telefon și în chatul de pe site, agenți de chat în Agent AI



Un host, multe pagini

Din URL-ul de start, crawlerul urmărește linkurile de pe același host, cu sau fără www. Sitemap-urile din robots.txt și un sitemap propriu extind lista.

Stabiliți domeniul

Pagini maxime și adâncime de link pe sursă, plus prefixe de cale, excluderi cu glob, URL-uri cu șir de interogare și căi de listă precum /tag.

JavaScript inclus

Fast citește HTML-ul, Full redă pagina într-un browser, Auto comută singur când o pagină pare goală. Un selector de clic deschide, de exemplu, acordeoanele FAQ.

Articole curate

Excluderea prin CSS ține în afara articolului zone precum navigarea sau subsolul. Dacă doriți, articolul primește un scurt rezumat din pagină.

Regulile site-ului

robots.txt și noindex sunt respectate. Dacă o pagină indexată este blocată ulterior, crawlerul își retrage articolul.

Modificări, nu rescrieri complete

Crawlerul recunoaște paginile modificate după conținut și lasă neatinse cele nemodificate. Paginile care dispar se retrag după două rulări fără a fi găsite.

02 · CUM FUNCȚIONEAZĂ

Adăugați o sursă, verificați-o, atribuiți-o agentului.

O sursă de pe site cere doar câteva câmpuri. Previzualizarea arată ce găsește crawlerul înainte de prima rulare, iar catalogul arată apoi ce s-a întâmplat cu fiecare URL.

1

Adăugați o sursă

Sub `Knowledge Base AI › Surse web`, folosiți `Adăugare sursă`: introduceți numele, URL-ul de start, baza de cunoștințe țintă și intervalul de actualizare.

2

Verificați previzualizarea

`Previzualizare` arată robots.txt, sitemap-urile și candidații, fără să indexeze nimic. Apoi setați domeniul și crawlerul.

3

Rulați și urmăriți

Prima rulare pornește la câteva minute după salvare, iar `Rulare acum` programează o altă rulare. `Pagini indexate` arată statusul și codul HTTP pentru fiecare URL.

4

Atribuiți-o agentului

Sub `Conversation AI › Agenți de conversație`, alegeți baza de cunoștințe din fila `Cunoștințe`. De atunci, agentul răspunde din articolele de pe site.

DATE TEHNICE

CARACTERISTICĂ	DETALII
Intrare	Pagini HTTPS publice ale unui singur host, text HTML
Domeniu pe sursă	Până la 2000 de pagini (implicit 200), adâncime de link până la 6 (implicit 3)
Surse pe cont	Până la 25
Crawler	Fast, Full sau Auto, cu selectori de clic și de așteptare
Ieșire	Articol intern pe pagină, cu URL-ul paginii, limba detectată și etichetele website și host
Prezentare generală	`Pagini indexate` cu status, redare, HTTP și ultimul fetch; `Istoricul sincronizării` pe sursă
Acces și facturare	Aplicația web sub `Knowledge Base AI › Surse web`; facturare după utilizare

03 · ÎN PRACTICĂ

Unde cunoștințele de pe site ajung la telefon și în chat.

Ghiduri pentru chatul de pe site

O companie de software își ține ghidurile sub /docs. O sursă de pe site cu prefixul de cale /docs le citește zilnic, iar chatul de pe site indică în răspuns pagina potrivită.

Pagini de produs pentru linia telefonică

Un comerciant își are paginile de produs citite săptămânal. Agentul conversațional de la telefon răspunde la întrebări despre mărimi și materiale din aceleași texte pe care le văd clienții online.

Site-ul plus articolele proprii

Tot ce nu se găsește pe site, de exemplu o închidere de sărbători, adăugați ca articol propriu în aceeași bază de cunoștințe. Agentul citește ambele împreună.

ÎNTREBĂRI FRECVENTE

Ce pagini citește o sursă de pe site?

Pagini HTTPS publice ale unui singur host, pornind de la URL-ul de start și, dacă doriți, prin sitemap-uri și alte URL-uri de start. Citește textul paginilor HTML, fără autentificare, cookie-uri sau credențiale.

Cât de actuale sunt cunoștințele agentului?

Atât de actuale cât este ultima rulare. Alegeți programul de la fiecare 6 ore până lunar, iar Rulare acum programează o rulare suplimentară. Două rulări manuale sunt la 15 minute distanță.

Ce se întâmplă cu modificările mele la un articol de pe site?

Când pagina se modifică, următoarea rulare rescrie articolul. Adăugirile proprii trebuie, așadar, puse într-un articol propriu, în aceeași bază de cunoștințe.

Apar paginile în portalul clienților?

Articolele de pe site sunt interne: echipa dumneavoastră și agenții bazei de cunoștințe le citesc, iar ele nu apar în portalul clienților.

ÎNCEPEȚI

Faceți din site-ul dumneavoastră prima sursă.

În aplicația web, mergeți la Knowledge Base AI › Surse web, folosiți Adăugare sursă pentru a introduce URL-ul de start, verificați previzualizarea și salvați. Prima rulare pornește câteva minute mai târziu.

paperoffice.ai/ro/ai-web-sources-site-web-cunostinte-agenti

MAI MULTE DESPRE MODUL

paperoffice.ai/ro/ai-web-sources-site-web-cunostinte-agenti

PREȚURI

paperoffice.ai/en/pricing

CONTACT

paperoffice.ai/ro/contact