



KARTA TECHNICZNA · PRZECHWYTYWANIE · ŹRÓDŁA ZE STRONY WWW

Wszystko jest na stronie. Kto to przepisuje?

Strona internetowa staje się wiedzą agenta.

Źródła ze strony czytają publiczne strony HTTPS witryny według harmonogramu. Każda strona staje się wewnętrznym artykułem w wybranej bazie wiedzy, a agenci odpowiadają na jej podstawie przez telefon i na czacie na stronie.

NA PIERWSZY RZUT OKA

Do 2000 stron na źródło, głębokość linków do 6, harmonogram od co 6 godzin do miesięcznie. Zmienione strony są przepisywane, a zniknięte są wycofywane po dwóch przebiegach.

01 · O CZYM TO JEST

Dwie wersje tej samej odpowiedzi. Teraz strona ustala fakty.

Godziny otwarcia, strony produktów i teksty pomocy od lat leżą na stronie. Ręczne przepisywanie ich dla agenta zajmuje czas, a każda zmiana na stronie sprawia, że kopia się dezaktualizuje. Źródła ze strony czytają same strony i przenoszą zmiany przy najbliższym przebiegu.

WEJŚCIE

Adres startowy z https, opcjonalna mapa witryny i kolejne adresy startowe na tym samym hoście

WYNIK

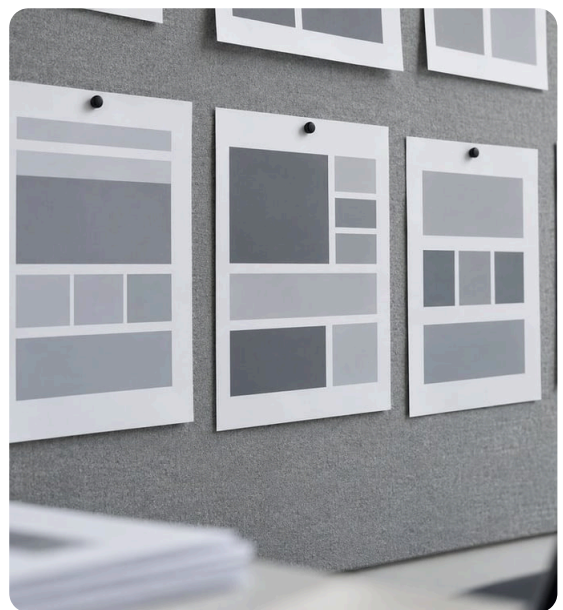
Jeden wewnętrzny artykuł na stronę, z adresem strony jako źródłem i wykrytym językiem

HARMONOGRAM

Co 6 lub 12 godzin, codziennie, co tydzień, co 14 dni lub miesięcznie

CZYTELNICY

Agenci konwersacji przez telefon i na czacie na stronie, agenci czatu w Agent AI

**Jeden host, wiele stron**

Od adresu startowego crawler podąża za linkami na tym samym hoście, z www lub bez. Mapy witryny z robots.txt oraz własna mapa rozszerzają listę.

Zakres

Maksymalna liczba stron i głębokość linków na źródło, a także prefiksy ścieżek, wzorce wykluczeń, adresy z parametrami zapytania oraz ścieżki list, np. /tag.

JavaScript w zestawie

Fast czyta HTML, Full renderuje w przeglądarce, Auto przetacza się samo, gdy strona wygląda na pustą. Selektor kliknięcia otwiera np. akordeony FAQ.

Czyste artykuły

Wykluczenie CSS trzyma obszary takie jak nawigacja czy stopka poza artykułem. W razie potrzeby artykuł dostaje krótkie streszczenie ze strony.

Reguły witryny

robots.txt i noindex są respektowane. Jeśli indeksowana strona zostanie później zablokowana, crawler wycofuje jej artykuł.

Zmiany, nie pełne przepisywanie

Crawler rozpoznaje zmienione strony po treści i nie rusza niezmienionych. Strony, które zniknęły, są wycofywane po dwóch przebiegach bez trafienia.

02 · JAK TO DZIAŁA

Dodać źródło, sprawdzić, przypisać je agentowi.

Źródło strony wymaga kilku pól. Podgląd pokazuje, co crawler znajduje przed pierwszym przebiegiem, a lista potem pokazuje, co stało się z każdym adresem URL.

1

Dodać źródło

W Knowledge Base AI > Źródła stron internetowych użyć Dodaj źródło : podać nazwę, adres startowy, docelową bazę wiedzy i interwał odświeżania.

2

Sprawdzić podgląd

Podgląd pokazuje robots.txt, mapy witryny i kandydatów bez indeksowania czegokolwiek. Dopiero potem ustawia się zakres i crawler.

3

Uruchomić i monitorować

Pierwszy przebieg startuje kilka minut po zapisaniu, a Przeskanuj teraz planuje kolejny. Zindeksowane strony pokazuje status i kod HTTP dla każdego adresu URL.

4

Przypisać agentowi

W Conversation AI > Agenci rozmów wybrać bazę wiedzy na karcie Wiedza . Od tej pory agent odpowiada na podstawie artykułów ze strony.

DANE TECHNICZNE

FUNKCJA	SZCZEGÓŁY
Wejście	Publiczne strony HTTPS jednego hosta, tekst HTML
Zakres na źródło	Do 2000 stron (domyślnie 200), głębokość linków do 6 (domyślnie 3)
Źródła na konto	Do 25
Crawler	Fast, Full lub Auto, z selektorami kliknięcia i oczekiwania
Wynik	Wewnętrzny artykuł na stronę z adresem URL, wykrytym językiem i tagami website oraz host
Przegląd	Zindeksowane strony ze statusem, renderowaniem, HTTP i ostatnim pobraniem; Historia synchronizacji dla każdego źródła
Dostęp i rozliczenie	Aplikacja webowa w Knowledge Base AI > Źródła stron internetowych ; rozliczenie według zużycia

03 · W PRAKTYCE

Gdzie wiedza ze strony dociera do telefonu i czatu.

Instrukcje dla czatu na stronie

Firma software trzyma instrukcje pod /docs. Źródło strony z prefiksem ścieżki /docs czyta je codziennie, a czat na stronie wskazuje pasującą stronę w odpowiedzi.

Strony produktów dla infolinii

Sklep detaliczny ma strony produktów czytane co tydzień. Agent konwersacji przez telefon odpowiada na pytania o rozmiary i materiały na podstawie tych samych tekstów, które klienci widzą w sieci.

Strona plus własne artykuły

Wszystko, czego nie ma na stronie, np. przerwa w pracy na święta, dodaje się jako własny artykuł w tej samej bazie wiedzy. Agent czyta oba źródła razem.

NAJCZĘŚCIEJ ZADAWANE PYTANIA

Jakie strony czyta źródło strony?

Publiczne strony HTTPS jednego hosta, od adresu startowego i, jeśli trzeba, przez mapy witryny oraz kolejne adresy startowe. Czyta tekst stron HTML, bez logowania, plików cookie i danych dostępowych.

Co się dzieje z moimi zmianami w artykule ze strony?

Gdy strona się zmieni, przy następnym przebiegu artykuł zostaje przepisany. Dopiski należy więc umieszczać w osobnym artykule w tej samej bazie wiedzy.

Jak aktualna jest wiedza agenta?

Tak aktualna, jak ostatni przebieg. Harmonogram wybiera się od co 6 godzin do miesięcznie, a [Przeskanuj teraz](#) planuje dodatkowy przebieg. Dwa ręczne przebiegi dzieli 15 minut.

Czy strony pojawiają się w portalu klienta?

Artykuły ze strony są wewnętrzne: czyta je zespół i agenci danej bazy wiedzy, a do portalu klienta nie trafiają.

ROZPOCZNIJ

Ustaw swoją stronę jako pierwsze źródło.

W aplikacji webowej przejdź do [Knowledge Base AI > Źródła stron internetowych](#), użyj

[Dodaj źródło](#), aby podać adres startowy, sprawdź podgląd i zapisz. Pierwszy przebieg startuje kilka minut później.

paperoffice.ai/pl/ai-web-sources-zrodlo-wiedzy-agregatora-stron-internetowych



WIĘCEJ O MODULE

paperoffice.ai/pl/ai-web-sources-zrodlo-wiedzy-agregatora-stron-internetowych

CENNIK

paperoffice.ai/en/pricing

KONTAKT

paperoffice.ai/pl/kontakt