



DATENBLATT · ERFASSEN · WEBSITE-QUELLEN

Alles steht auf der Website. Wer tippt es ab?

Ihre Website wird zum Wissen Ihres Agenten.

Website-Quellen lesen die öffentlichen HTTPS-Seiten Ihrer Website nach Zeitplan ein. Jede Seite wird ein interner Artikel in der Wissensdatenbank, die Sie wählen, und Ihre Agenten antworten daraus am Telefon und im Website-Chat.

AUF EINEN BLICK

Bis zu 2000 Seiten je Quelle, Linktiefe bis 6, Rhythmus von alle 6 Stunden bis monatlich. Geänderte Seiten werden neu geschrieben, verschwundene nach zwei Läufen zurückgezogen.

01 · WORUM ES GEHT

Zwei Stände derselben Antwort. Ab jetzt zählt Ihre Website.

Öffnungszeiten, Produktseiten und Hilfetexte stehen längst auf Ihrer Website. Sie für den Agenten abzutippen kostet Zeit, und jede Änderung auf der Website macht die Kopie alt. Website-Quellen lesen die Seiten selbst und bringen Änderungen mit dem nächsten Lauf mit.

EINGABE

Start-URL mit https, optional Sitemap und weitere Start-URLs auf demselben Host

ERGEBNIS

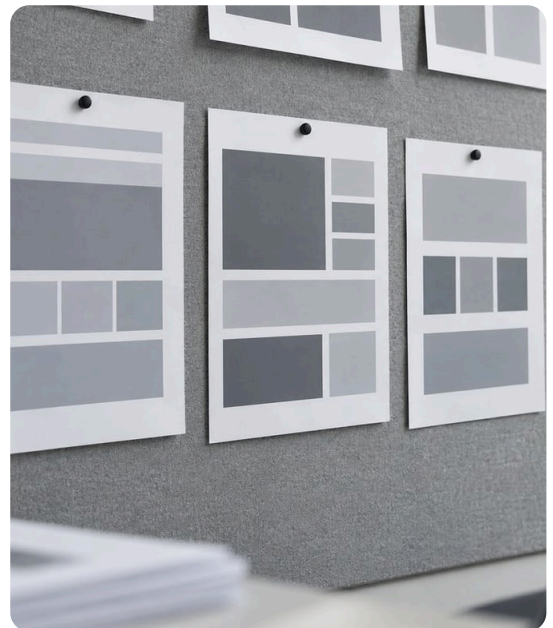
Ein interner Artikel je Seite, mit der Seiten-URL als Quelle und erkannter Sprache

RHYTHMUS

Alle 6 oder 12 Stunden, täglich, wöchentlich, alle 14 Tage oder monatlich

LESER

Konversationsagenten am Telefon und im Website-Chat, Chat-Agenten in Agent AI

**Ein Host, viele Seiten**

Ab der Start-URL folgt der Crawler den Links auf demselben Host, wahlweise mit www. Sitemaps aus robots.txt und eine eigene Sitemap ergänzen die Liste.

Umfang festlegen

Höchstzahl der Seiten und Linktiefe je Quelle, dazu Pfad-Prefixe, Ausschluss-Globs, URLs mit Query-String und Listen-Pfade wie /tag.

Auch mit JavaScript

Fast liest das HTML, Full rendert im Browser, Auto wechselt selbst, wenn eine Seite leer wirkt. Ein Klick-Selektor öffnet etwa FAQ-Akkordeons.

Saubere Artikel

Der CSS-Ausschluss hält Bereiche wie Navigation oder Fußzeile aus dem Artikel. Auf Wunsch übernimmt der Artikel einen Kurztext aus der Seite.

Regeln der Website

robots.txt und noindex werden beachtet. Wird eine indexierte Seite später gesperrt, zieht der Crawler ihren Artikel zurück.

Änderungen statt Komplettlauf

Geänderte Seiten erkennt der Crawler am Inhalt, unveränderte bleiben stehen. Verschwundene Seiten werden nach zwei Läufen ohne Treffer zurückgezogen.

02 · SO FUNKTIONIERT ES

Quelle anlegen, prüfen, dem Agenten zuweisen.

Eine Website-Quelle ist in wenigen Feldern angelegt. Die Vorschau zeigt vor dem ersten Lauf, was der Crawler findet, und der Katalog zeigt danach je URL, was passiert ist.

1

Quelle anlegen

Unter **Knowledge Base AI** › **Website-Quellen** mit **Quelle anlegen** : Name, Start-URL, Ziel-Wissensdatenbank und Rhythmus der Aktualisierung eintragen.

2

Vorschau prüfen

Die **Vorschau** zeigt robots.txt, Sitemaps und Kandidaten, ohne etwas zu indexieren. Danach Umfang und Crawler einstellen.

3

Lauf starten und kontrollieren

Der erste Lauf startet wenige Minuten nach dem Speichern, **Jetzt crawlen** plant einen weiteren ein. **Indexierte Seiten** zeigt je URL Status und HTTP-Code.

4

Dem Agenten zuweisen

Unter **Conversation AI** › **Konversationsagenten** im Reiter **Wissen** die Wissensdatenbank wählen. Ab dann antwortet der Agent aus den Website-Artikeln.

TECHNISCHE DATEN

MERKMAL	DETAILS
Eingabe	Öffentliche HTTPS-Seiten eines Hosts, HTML-Text
Umfang je Quelle	Bis zu 2000 Seiten (voreingestellt 200), Linktiefe bis 6 (voreingestellt 3)
Quellen je Konto	Bis zu 25
Crawler	Fast, Full oder Auto, mit Klick- und Warte-Selektor
Ausgabe	Interner Artikel je Seite mit Seiten-URL, erkannter Sprache und den Tags website und Host
Überblick	Indexierte Seiten mit Status, Rendern, HTTP und letztem Abruf; Sync-Historie je Quelle
Zugang und Abrechnung	Web-App unter Knowledge Base AI › Website-Quellen ; Abrechnung nach Verbrauch

03 · IM EINSATZ

Wo Website-Wissen am Telefon und im Chat ankommt.

Anleitungen für den Website-Chat

Ein Softwarehaus pflegt seine Anleitungen unter /docs. Eine Website-Quelle mit dem Pfad-Prefix /docs liest sie täglich, und der Website-Chat nennt in seiner Antwort die passende Seite.

Produktseiten für die Hotline

Ein Händler lässt seine Produktseiten wöchentlich lesen. Der Konversationsagent am Telefon beantwortet Fragen zu Maßen und Material aus denselben Texten, die Kunden online sehen.

Website plus eigene Artikel

Was nicht auf der Website steht, etwa Betriebsferien, ergänzen Sie als eigenen Artikel in derselben Wissensdatenbank. Der Agent liest beides zusammen.

HÄUFIGE FRAGEN

Welche Seiten liest eine Website-Quelle?

Öffentliche HTTPS-Seiten eines Hosts, ab der Start-URL und auf Wunsch über Sitemaps und weitere Start-URLs. Gelesen wird der Text der HTML-Seiten, ohne Login, Cookies oder Zugangsdaten.

Was passiert mit eigenen Änderungen an einem Website-Artikel?

Ändert sich die Seite, schreibt der nächste Lauf den Artikel neu. Ergänzungen gehören deshalb in einen eigenen Artikel derselben Wissensdatenbank.

Wie aktuell ist das Wissen des Agenten?

So aktuell wie der letzte Lauf. Den Rhythmus wählen Sie von alle 6 Stunden bis monatlich, Jetzt crawlen plant einen zusätzlichen Lauf ein. Zwischen zwei manuellen Läufen liegen 15 Minuten.

Erscheinen die Seiten im Kundenportal?

Website-Artikel sind intern: Ihr Team und die Agenten der Wissensdatenbank lesen sie, im Kundenportal erscheinen sie nicht.

JETZT STARTEN

Machen Sie Ihre Website zur ersten Quelle.

In der Web-App unter [Knowledge Base AI > Website-Quellen](#) mit [Quelle anlegen](#) die Start-URL eintragen, die Vorschau prüfen und speichern. Der erste Lauf startet wenige Minuten später.

paperoffice.ai/de/ki-web-sources-website-crawler-wissensquelle



MEHR ZUM MODUL

paperoffice.ai/de/ki-web-sources-website-crawler-wissensquelle

PREISE

paperoffice.ai/de/preise

KONTAKT

paperoffice.ai/de/kontakt