



ТЕХНИЧЕСКОЕ ОПИСАНИЕ · ЗАХВАТ · ВЕБ-ИСТОЧНИКИ

Всё есть на вашем сайте. Кто это перепечатывает? Ваш сайт становится знанием вашего агента.

Веб-источники по расписанию читают публичные HTTPS-страницы вашего сайта. Каждая страница становится внутренней статьёй в выбранной вами базе знаний, а ваши агенты отвечают по ней по телефону и в чате на сайте.

КОРОТКО

До 2000 страниц на источник, глубина ссылок до 6, расписание от каждых 6 часов до ежемесячного. Изменённые страницы переписываются, исчезнувшие снимаются после двух запусков.

01 · О ЧЁМ РЕЧЬ

Две версии одного ответа. Теперь главным становится ваш сайт.

Часы работы, страницы товаров и тексты справки давно есть на вашем сайте. Перепечатывать их для агента отнимает время, а каждое изменение на сайте делает копию устаревшей. Веб-источники читают сами страницы и приносят изменения со следующим запуском.

ВХОД

Стартовый URL с https, по желанию карта сайта и другие стартовые URL на том же хосте

РЕЗУЛЬТАТ

Одна внутренняя статья на страницу, с URL страницы как источником и определённым языком

РАСПИСАНИЕ

Каждые 6 или 12 часов, ежедневно, еженедельно, каждые 14 дней или ежемесячно

ЧИТАТЕЛИ

Агенты диалога по телефону и в чате на сайте, чат-агенты в Agent AI

**Один хост, много страниц**

От стартового URL краулер проходит по ссылкам на том же хосте, с www или без него. Карты сайта из robots.txt и собственная карта сайта расширяют список.

Задайте охват

Максимум страниц и глубина ссылок на источник, а также префиксы путей, исключающие шаблоны, URL с параметрами запроса и списочные пути, например /tag.

JavaScript включён

Fast читает HTML, Full отрисовывает страницу в браузере, Auto переключается сам, когда страница выглядит пустой. Селектор клика открывает, например, раскрывающиеся блоки FAQ.

Чистые статьи

Исключение через CSS оставляет вне статьи такие области, как навигация или подвал. По желанию статья получает короткое краткое описание со страницы.

Правила сайта

Учитываются robots.txt и noindex. Если проиндексированная страница позже оказывается закрытой, краулер снимает её статью.

Изменения, а не полная перезапись

Краулер узнаёт изменённые страницы по содержимому и не трогает неизменённые. Исчезнувшие страницы снимаются после двух запусков без обнаружения.

02 · КАК ЭТО РАБОТАЕТ

Добавьте источник, проверьте, назначьте агенту.

Веб-источнику нужно всего несколько полей. Предпросмотр показывает, что найдёт краулер до первого запуска, а каталог затем показывает, что произошло с каждым URL.

1

Добавьте источник

В разделе Knowledge Base AI › Источники веб-сайтов используйте Добавить источник : введите название, стартовый URL, целевую базу знаний и интервал обновления.

2

Проверьте предпросмотр

Предпросмотр показывает robots.txt, карты сайта и кандидатов без индексации чего-либо. Затем задайте охват и краулер.

3

Запустите и следите

Первый запуск стартует через несколько минут после сохранения, а Запустить сканирование сейчас ставит ещё один. Индексированные страницы показывает статус и HTTP-код для каждого URL.

4

Назначьте агенту

В разделе Conversation AI › Агенты разговоров выберите базу знаний на вкладке Знания . С этого момента агент отвечает по статьям с сайта.

ТЕХНИЧЕСКИЕ ДАННЫЕ

ФУНКЦИЯ	ПОДРОБНОСТИ
Вход	Публичные HTTPS-страницы одного хоста, текст HTML
Охват на источник	До 2000 страниц (по умолчанию 200), глубина ссылок до 6 (по умолчанию 3)
Источников на аккаунт	До 25
Краулер	Fast, Full или Auto, с селекторами клика и ожидания
Вывод	Внутренняя статья на страницу с URL страницы, определённым языком и тегами website и host
Обзор	Индексированные страницы со статусом, рендерингом, HTTP и датой последней загрузки; История синхронизации по каждому источнику
Доступ и оплата	Веб-приложение в разделе Knowledge Base AI › Источники веб-сайтов ; оплата по использованию

03 · НА ПРАКТИКЕ

Там, где знания с сайта доходят до телефона и чата.

Инструкции для чата на сайте

Разработческая компания хранит свои инструкции под /docs. Веб-источник с префиксом пути /docs читает их ежедневно, а чат на сайте называет подходящую страницу в ответе.

Страницы товаров для горячей линии

Розничный продавец читает страницы товаров раз в неделю. Агент диалога по телефону отвечает на вопросы о размерах и материалах по тем же текстам, которые клиенты видят в интернете.

Сайт плюс собственные статьи

Всё, чего нет на сайте, например закрытие на праздники, вы добавляете как собственную статью в той же базе знаний. Агент читает оба источника вместе.

ЧАСТО ЗАДАВАЕМЫЕ ВОПРОСЫ

Какие страницы читает веб-источник?

Публичные HTTPS-страницы одного хоста, начиная со стартового URL и, если хотите, через карты сайта и дополнительные стартовые URL. Он читает текст HTML-страниц, без входа, cookie или учётных данных.

Что происходит с моими правками статьи с сайта?

Когда страница меняется, при следующем запуске статья переписывается. Дополнения поэтому нужно вести в отдельной статье в той же базе знаний.

Насколько актуальны знания агента?

Настолько, насколько актуален последний запуск. Расписание вы выбираете от каждые 6 часов до ежемесячного, а Запустить сканирование сейчас ставит дополнительный запуск. Два запуска вручную идут с интервалом 15 минут.

Появляются ли страницы в клиентском портале?

Статьи с сайта внутренние: их читают ваша команда и агенты базы знаний, в клиентский портал они не попадают.

НАЧАТЬ РАБОТУ

Сделайте свой сайт первым источником.

В веб-приложении откройте Knowledge Base AI › Источники веб-сайтов, с помощью Добавить источник введите стартовый URL, проверьте предпросмотр и сохраните. Первый запуск начнётся через несколько минут.

paperoffice.ai/ru/ai-web-sources-agent-istochnik-znaniy-veb-kraulera



ПОДРОБНЕЕ О МОДУЛЕ

paperoffice.ai/ru/ai-web-sources-agent-istochnik-znaniy-veb-kraulera

ЦЕНЫ

paperoffice.ai/ru/tseny

КОНТАКТЫ

paperoffice.ai/ru/kontakt