
ItkRobotsTxt

Anwenderdokumentation

André Fischer

30.03.2026

Inhaltsverzeichnis

1 Was macht dieses Plugin?	2
1.1 Diagramm: Wo der Text herkommt	2
2 Warum sich das lohnt	2
3 Wo du die Einstellung findest	3
4 Plugin-Konfiguration (technisch)	3
5 Typischer Ablauf im Alltag	3
6 Beispielinhalt	4
7 Fallback, wenn das Feld leer ist	4
8 Unterschiedliche Inhalte pro Sales Channel	4
9 Caching und „Ich sehe die Änderung nicht“	5
10 Grenzen des Plugins	5
11 Troubleshooting	5
12 FAQ	6

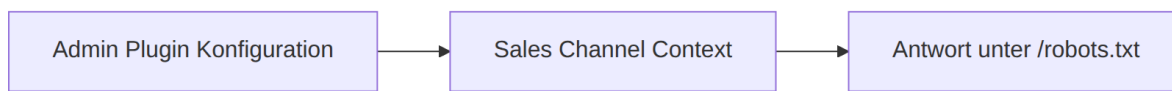
1 Was macht dieses Plugin?

Unter der festen Adresse `/robots.txt` erwarten Suchmaschinen und andere Crawler eine **kleine Textdatei** mit Regeln: welche Bereiche deines Shops sie besuchen dürfen und welche nicht. Ohne eigenes Plugin liefert Shopware eine Standard-`robots.txt`. Mit **ItkRobotsTxt** pflegst du den **kompletten Inhalt** dieser Datei **im Admin** – und kannst für jeden **Verkaufskanal (Sales Channel)** einen **eigenen Text** hinterlegen, wenn deine Domains unterschiedliche Anforderungen haben.

Das Plugin ersetzt nicht die Shopware-Sitemap-Generierung, sondern nur **das, was unter `/robots.txt` steht**. Alles andere (SEO-URLs, Meta-Robots an Seiten, HTTP-Header) bleibt bei Shopware.

Wichtig: Eine bereits vorhandene `robots.txt` welche unter `/public` liegt muss entfernt werden. Die Datei hat eine höhere Priorität als der Controller und würde somit den Controller überschreiben.

1.1 Diagramm: Wo der Text herkommt



2 Warum sich das lohnt

Crawl-Budget und Lärm vermeiden: Wenn Bots tausende Varianten-URLs, interne Suche oder Konto-Seiten crawlen, verschwenden sie Kapazität und können Dubletten oder unnötige Indexeinträge erzeugen. In der `robots.txt` sagst du: „Diese Pfade bitte nicht.“

Staging und Kopien: Auf Testshops willst du oft **alles sperren (Disallow: /)**, damit nichts in den Index rutscht. Mit kanalspezifischer Konfiguration kann der Live-Kanal offen bleiben und der Staging-Kanal zu.

Sitemap-Hinweis: Viele tragen eine Zeile **Sitemap: `https://...`** ein, damit Crawler die Sitemap finden. Die Sitemap selbst erzeugt Shopware woanders – hier geht es nur um den **Verweis**.

Wichtig zu verstehen: Die `robots.txt` ist eine **Höflichkeitsregel**. Sie ist **kein Zugriffsschutz**: Ein Angreifer oder ein ignoranter Bot kann trotzdem URLs aufrufen. Für echte Absicherung brauchst du Auth, Firewall, Admin hinter VPN usw.

3 Wo du die Einstellung findest

1. Im Shopware-Admin: **Erweiterungen** → **Meine Erweiterungen**
2. Plugin „**Robots.txt pro Sales Channel**“ / **ItkRobotsTxt** öffnen und auf **Konfiguration** klicken.
3. Oben den **Sales Channel** wählen, für den der Text gelten soll.

4 Plugin-Konfiguration (technisch)

Anzeige im Admin	Interner Config-Key	Typ
Inhalt der robots.txt	robotsTxtContent	Textarea (mehrzeilig)

Vollständiger SystemConfig-Pfad im Code: `ItkRobotsTxt.config.robotsTxtContent`.

Der Text wird **unverändert** ausgegeben: kein automatisches Escaping, keine Prüfung auf „gültige“ robots-Syntax. Du bist dafür verantwortlich, Tippfehler und Zeilenumbrüche (`\\n`) sinnvoll zu setzen.

5 Typischer Ablauf im Alltag

1. Legt fest, welche Pfade **nicht** gecrawlt werden sollen (Konto, Checkout, interne Suche, Tracking-Parameter-Seiten ...).
2. Tragt den fertigen Block in das Textfeld ein und speicherst.
3. Prüfe im Browser <https://deine-domain.de/robots.txt> (Domain muss zum gewählten Sales Channel passen).
4. Bei mehreren Sprach- oder Marken-Domains: für **jeden Kanal** den passenden Inhalt eintragen (inkl. korrekter **Sitemap-URL** pro Domain).

6 Beispielinhalt

```
1 User-agent: *
2 Disallow: /admin
3 Disallow: /account/
4 Disallow: /checkout/
5 Disallow: /widgets/
6 Disallow: /search
7 Allow: /
8
9 Sitemap: https://dein-shop.de/sitemap.xml
```

Kurz erklärt:

- **User-agent:** * – die folgenden Regeln gelten für „alle“ Bots, die das respektieren.
- **Disallow:** ... – dieser Pfad soll nicht gecrawlt werden (Präfix-Logik wie in der robots-Spezifikation beachten).
- **Allow:** / – explizit erlauben, falls du vorher etwas eingeschränkt hast (Reihenfolge und Interpretation hängen vom Bot ab; bei Unsicherheit SEO-Quellen konsultieren).
- **Sitemap:** – vollständige URL zur Sitemap.

7 Fallback, wenn das Feld leer ist

Ist **kein** Text gespeichert (leerer String), liefert das Plugin automatisch:

```
1 User-agent: *
2 Allow: /
```

Damit verhält sich der Shop **offen** gegenüber Crawlern – für Produktion okay, für Staging oft **zu offen**. Leere Konfiguration ist also bewusst „permittierend“.

8 Unterschiedliche Inhalte pro Sales Channel

Jeder Verkaufskanal hat eigene Domains. Das Plugin liest die Konfiguration mit der **Sales-Channel-ID** des aktuellen Kontexts. Wenn du für Kanal A und Kanal B unterschiedliche Texte brauchst, wechselst

du im Konfigurationsdialog den Kanal, speicherst jeweils den Inhalt – und unter jeder zugehörigen Domain siehst du die passende `robots.txt`.

Falls für einen Kanal **nichts** eingetragen ist, greift Shopwares üblicher **Fallback** der System-Config (globaler Standard), bis wieder ein leerer Wert zum oben genannten Minimal-Fallback führt – im Zweifel nach dem Speichern live testen.

9 Caching und „Ich sehe die Änderung nicht“

Die Antwort kommt mit **Cache-Control: public, max-age=3600** – Browser und zwischengeschaltete Caches dürfen die Datei bis zu **einer Stunde** wiederverwenden. Nach einer Änderung: harter Reload (**Strg+Shift+R** / **Cmd+Shift+R**) oder anderes Netzwerk testen. Falls immer noch nicht, sollte ein Administrator auf dem Server die **robots.txt unter public/** entfernen

10 Grenzen des Plugins

- **Keine** pro-URL-Logik aus der Datenbank – nur der Text, den du einträgst.
- **Keine** automatische Einbindung der Sitemap-URL (außer du schreibst sie selbst in den Text).
- **Kein** Ersatz für **noindex** in Meta-Tags oder für HTTP-Status – manche Seiten willst du zusätzlich oder stattdessen dort absichern.
- Die `robots.txt` **verhindert kein Crawling** durch böswillige Clients.

11 Troubleshooting

Symptom	Was du prüfst
Alte robots.txt trotz Speichern	Browser-Cache, CDN, max-age=3600 , public/robots.txt existiert?
Unter Domain A falscher Inhalt	Welcher Sales Channel gehört zu Domain A? Config für diesen Kanal prüfen.

Symptom	Was du prüfst
Gar kein Plugin-Inhalt / Shopware-Default	Ist das Plugin aktiv? Wird eine andere Route /robots.txt von einem anderen Bundle bedient? (Dann Entwickler fragen – Route-Konflikt.)
Zeilenumbrüche kaputt	Textarea: jede Direktive sinnvoll eine Zeile ; kein HTML einfügen.

12 FAQ

Soll ich /admin sperren?

Ja, das ist üblich – der Admin ist ohnehin login-geschützt, aber Crawler sollen ihn nicht einmal als URL „kennlernen“ müssen.

Ist Disallow genug für vertrauliche Bereiche?

Nein. Nutze Authentifizierung und Netzwerkregeln.

Wo stelle ich die Shopware-Sitemap ein?

In den **Shopware-Einstellungen** zur Sitemap, nicht in diesem Plugin. Hier trägst du nur die **Zeile Sitemap: ...** ein, wenn du willst.