

robots.txt verstehen und richtig einsetzen

PRAXISLEITFADEN | Stand: 27. Oktober 2026 | Kostenlos

1. Zweck verstehen

Vorgehen: robots.txt steuert Crawling-Regeln für kooperative Bots, ist aber kein Zugriffsschutz und garantiert keine Deindexierung.

Beispiel: <https://example.org/robots.txt>

Prüfkriterium: Datei ist öffentlich erreichbar und syntaktisch korrekt.

Vermeiden: Vertrauliche Pfade durch robots.txt schützen wollen.

2. Regeln definieren

Vorgehen: Nutze User-agent, Disallow und Allow gezielt; prüfe die Semantik des jeweiligen Crawlers.

Beispiel: User-agent: *

Disallow: /intern/

Allow: /

Prüfkriterium: Wichtige öffentliche Inhalte sind nicht versehentlich gesperrt.

Vermeiden: Disallow: / auf Produktion stehen lassen.

3. Sitemap referenzieren

Vorgehen: Verweise optional auf die XML-Sitemap mit absoluter URL.

Beispiel: Sitemap: <https://example.org/sitemap.xml>

Prüfkriterium: Sitemap lässt sich abrufen und verarbeiten.

Vermeiden: Sitemap-Pfad mit einer Sperrregel verwechseln.

4. Noindex getrennt behandeln

Vorgehen: Verwende für indexierbare abrufbare Seiten bei Bedarf robots-Meta noindex oder X-Robots-Tag; gesperrte Seiten können von Bots nicht ausgelesen werden.

Beispiel: `<meta name="robots" content="noindex,follow">`

Prüfkriterium: Crawler kann die Noindex-Anweisung abrufen.

Vermeiden: Disallow und noindex kombinieren und sofortige Entfernung erwarten.

5. Deployment testen

Vorgehen: Prüfe Regeln auf Live-Host und Staging getrennt. Teste mit URL-Prüfung und HTTP-Abruf.

Beispiel: `curl https://example.org/robots.txt`

Prüfkriterium: Produktionsregeln stimmen mit dem gewünschten Crawling überein.

Vermeiden: Staging-Robots-Datei unbemerkt übernehmen.

Fachquellen

Google: robots.txt – <https://developers.google.com/search/docs/crawling-indexing/robots/intro>

Onlineversion: <https://www.cyberknecht.de/wissen/anleitungen/robots-txt-anleitung.html>