

# Co je soubor robots.txt

Poslední aktualizace 29 listopadu, 2024

Robots.txt je jednoduchý textový soubor, který slouží k řízení přístupu vyhledávacích robotů (tzv. crawlerů) na váš web.

Nachází se v kořenovém adresáři webu (např. <https://webglobe.cz/robots.txt>) a obsahuje pokyny, které určují, které části webu mohou nebo nemohou roboti indexovat.

Tento soubor je důležitým nástrojem pro správu indexace a pomáhá optimalizovat způsob, jakým vyhledávače, jako Google, Seznam nebo Bing, procházejí váš web.

## Kdy se robots.txt upravuje?

- Omezení přístupu:** Přes robots.txt můžete robotům zakázat přístup k citlivým částem webu (např. administrace, testovací stránky nebo duplicitní obsah).
- Optimalizace indexace:** Robots.txt řídí, na které stránky nebo adresáře se mají vyhledávače zaměřit, čímž šetří kapacitu procházení (tzv. crawl budget).
- Pokyny pro specifické roboty:** Robots.txt umožňuje zadat pravidla pro konkrétní vyhledávače nebo roboty.

## Příklad souboru robots.txt

*User-agent: \**

*Disallow: /admin/*

*Disallow: /private-data/*

*Sitemap: <https://priklad.cz/sitemap.xml>*

- **User-agent: \*** znamená, že pravidla platí pro všechny roboty.
- **Disallow:** Zakazuje přístup do složek `/admin/` a `/private-data/`.
- **Sitemap:** Umožňuje robotům snadno najít XML sitemapu, která obsahuje seznam stránek určených k indexaci.

## Co robots.txt nedokáže?

Robots.txt **nechrání data před přístupem** – zakázané stránky mohou být stále přístupné pro uživatele, kteří znají přímý odkaz.

Neposkytuje také 100% garanci, že všichni roboti budou pravidla respektovat. Nezodpovědní roboti mohou soubor ignorovat.

## **Proč je robots.txt důležitý?**

Správné nastavení souboru robots.txt je důležité pro optimalizaci pro vyhledávače (SEO).

Pomáhá zabránit tomu, aby byly indexovány zbytečné nebo nežádoucí stránky, čímž se zlepšuje kvalita indexace vašeho webu.

Je to jednoduchý, ale účinný nástroj pro řízení chování vyhledávacích robotů.

*Přečtěte si, jak vytvořit sitemapu svého webu.*