**Chcesz ograniczyć robotom dostęp do Twojej strony? Użyj pliku robots.txt! 🤖**

**Chcesz ograniczyć robotom dostęp do Twojej strony? Użyj pliku robots.txt! 🤖**

**Robots.txt to plik tekstowy umieszczany w głównym katalogu domeny pod adresem nazwadomeny.tld/robots.txt. Służy do blokowania robotom dostępu do crawlowania wybranych zasobów strony. Co ważne, przy jego pomocy możesz nie tylko zablokować możliwość czytania wybranych podstron Googlebotowi, ale także crawlerom takich programów jak Ahrefs, czy Semrush, a także różnego rodzaju spamerskim robotom obciążającym zasoby serwera. Pamiętaj, że jeżeli całkowicie zablokujesz dostęp do serwisu robotom Google, Twoja strona może mieć problemy z indeksowaniem. Jednocześnie należy pamiętać, że funkcją robots.txt jest sterowanie crawlowaniem, a nie stricte indeksacją zasobów.**

**W celu zablokowania indeksacji stosuje się dyrektywę noindex w tagu meta robots lub w nagłówkach HTTP. W szczególnych przypadkach Google sugeruje, aby nawet stosować zabezpieczenie plików hasłem na serwerze. Blokowanie botów ahrefs i innych tego typu programów, stosuje się często na zapleczach, aby nie były zaczytywane dane o linkach wychodzących. Warto jednak wziąć pod uwagę blokowanie tych botów z poziomu pliku .htaccess, a nie robots.txt, aby nie tworzyć niepotrzebnych footprintów przy większej ilości stron zapleczowych.**

**W przypadku blokowania robotów mamy dwie możliwości:**

**🟥 zablokowanie dostępu do całego serwisu,**

**🟥 zablokowanie dostępu jedynie do wybranych części serwisu (np. całych katalogów),**

**🟥 zablokowanie dostępu do konkretnych formatów lub pojedynczych plików.**

**Po co właściwie blokuje się dostęp do serwisu❓**

**Blokując crawlowanie niepożądanych zasobów możesz zoptymalizować crawl budget. Dzięki temu Googlebot będzie efektywniej znajdował strony, na których Ci zależy.**

**Natomiast jeśli chodzi o blokowanie narzędzi, jest to pewnego rodzaju zabezpieczenie przed konkurentami, którzy chcieliby Twój serwis. Dodatkowo ukryjesz linki wychodzące z PBN-ów, co pomoże ukryć Twoją strategię linkowania. 🔗**

**Najczęściej blokuje się strony, które nie są ważne z punktu widzenia wyszukiwania. Są nimi np. strony logowania, katalogi ze skryptami, wtyczkami i innymi nieistotnymi dla SEO danymi.**

**Jeśli planujesz pobawić się robots.txt warto zapoznać się z całą dokumentacją dotyczącą tego pliku. W komentarzu znajdziesz do niej link.**

**#robotstxt #seo #pozycjonowanie**

Chcesz ograniczyć robotom dostęp do Twojej strony? Użyj pliku robots.txt! 🤖

Robots.txt to plik tekstowy umieszczany w głównym katalogu domeny pod adresem nazwadomeny.tld/robots.txt. Służy do blokowania robotom dostępu do crawlowania wybranych zasobów strony. Co ważne, przy jego pomocy możesz nie tylko zablokować możliwość czytania wybranych podstron Googlebotowi, ale także crawlerom takich programów jak Ahrefs, czy Semrush, a także różnego rodzaju spamerskim robotom obciążającym zasoby serwera. Pamiętaj, że jeżeli całkowicie zablokujesz dostęp do serwisu robotom Google, Twoja strona może mieć problemy z indeksowaniem. Jednocześnie należy pamiętać, że funkcją robots.txt jest sterowanie crawlowaniem, a nie stricte indeksacją zasobów.

W celu zablokowania indeksacji stosuje się dyrektywę noindex w tagu meta robots lub w nagłówkach HTTP. W szczególnych przypadkach Google sugeruje, aby nawet stosować zabezpieczenie plików hasłem na serwerze. Blokowanie botów ahrefs i innych tego typu programów, stosuje się często na zapleczach, aby nie były zaczytywane dane o linkach wychodzących. Warto jednak wziąć pod uwagę blokowanie tych botów z poziomu pliku .htaccess, a nie robots.txt, aby nie tworzyć niepotrzebnych footprintów przy większej ilości stron zapleczowych.

W przypadku blokowania robotów mamy dwie możliwości:

🟥 zablokowanie dostępu do całego serwisu,

🟥 zablokowanie dostępu jedynie do wybranych części serwisu (np. całych katalogów),

🟥 zablokowanie dostępu do konkretnych formatów lub pojedynczych plików.

Po co właściwie blokuje się dostęp do serwisu❓

Blokując crawlowanie niepożądanych zasobów możesz zoptymalizować crawl budget. Dzięki temu Googlebot będzie efektywniej znajdował strony, na których Ci zależy.

Natomiast jeśli chodzi o blokowanie narzędzi, jest to pewnego rodzaju zabezpieczenie przed konkurentami, którzy chcieliby Twój serwis. Dodatkowo ukryjesz linki wychodzące z PBN-ów, co pomoże ukryć Twoją strategię linkowania. 🔗

Najczęściej blokuje się strony, które nie są ważne z punktu widzenia wyszukiwania. Są nimi np. strony logowania, katalogi ze skryptami, wtyczkami i innymi nieistotnymi dla SEO danymi.

Jeśli planujesz pobawić się robots.txt warto zapoznać się z całą dokumentacją dotyczącą tego pliku. W komentarzu znajdziesz do niej link.

#robotstxt #seo #pozycjonowanie