Jak ograniczyć ruch robotów na[..]

2024-04-24 01:11:48

Wydruk artykułu FAQ

Kategoria:	Hosting ITL::Inne	**Ostatnia aktualizacja:**	2017-06-07 12:50:41
Status:	publiczny (dla wszystkich)		
Język:	pl		

Słowa kluczowe

transfer robots.txt

Objawy (publiczny)

Na mojej stronie internetowej dużo ruchu generują nieznane roboty internetowe.

Problem (publiczny)

W internecie różne firmy wspierające marketing internetowy, zbierają informacje o stronach WWW, wykorzystując do tego roboty internetowe. Czasem ruch tych robotów mocno obciąża stronę, jednocześnie nie przynosząc właścicielowi strony żadnych korzyści.

Rozwiązanie (publiczny)

Można ograniczyć ruch robotów internetowych na stronie, odpowiednio przygotowując plik robots.txt.
Plik należy umieścić w głównym katalogu każdego serwisu www, tak by był dostępny w następujący sposób:
http://moja.domena.pl/robots.txt

Przykładowa treść pliku robots.txt, blokująca wszystkie roboty oprócz robotów wyszukiwarek Google i Microsoft Bing:

```
User-agent: googlebot
Disallow:

User-agent: bingbot
Disallow:

User-agent: *
Disallow: /
```

Uwaga: roboty internetowe traktują zawartość pliku robots.txt jako wskazówki. Większość z robotów do tych wskazówek się stosuje. Jednak może się zdarzyć, że robot celowo i natarczywie ignoruje zawartość pliku robots.txt. Wówczas można takiego robota zablokować mniej przyjazny (dla robota) sposób: [1]w pliku .htaccess

[1] https://support.intertele.pl/otrs/public.pl?Action=PublicFAQZoom;ItemID=21