

Boti a crawlery: keď váš web spomaľujú návštevníci, ktorí nikdy nič nekúpia

21.8.2026 - Jakub Krakovík | Websupport

Máte e-shop a stránka sa vám zrazu začne načítavať pomaly. Pozriete sa na návštevnosť a nič zvláštne sa nedeje. Ľudí na webe máte možno dokonca menej ako minulý týždeň. Napriek tomu web reaguje pomaly. Občas seká. Občas sa nenačíta vôbec. Kto ho teda navštevuje? Možno nie ľudia, ale boti.

Čo je bot a prečo vôbec chodí na váš web?

Bot je program, ktorý na internete automaticky vykonáva určité úlohy. Crawler je typ bota, ktorý prechádza webové stránky, sleduje odkazy a načítava ich obsah.

Niektoré crawlery sú pre váš web užitočné. Napríklad Googlebot pomáha vyhľadávaču objaviť a spracovať váš obsah, aby sa mohol zobrazovať vo výsledkoch vyhľadávania. Podobne fungujú aj crawlery pre AI vyhľadávanie či nástroje na monitoring a SEO analýzu.

Iné boty môžu web zbytočne zaťažovať alebo mať škodlivé úmysly. **Bot teda nie je automaticky problém. Rozhoduje, čo na vašom webe robí a koľko jeho zdrojov pritom spotrebuje.**

Boti pritom nie sú okrajová záležitosť. Podľa Thales Bad Bot Report, ktorý analyzuje prevádzku za rok 2025, automatizované požiadavky tvorili približne **53 % webovej prevádzky**. Človek už teda nie je jediný (a ani väčšinový) návštevník internetu.

V čom je rozdiel medzi dobrými a zlými botmi?

Užitoční boti

Pomáhajú vášmu webu fungovať alebo byť viditeľný. Patria sem napríklad crawlery vyhľadávačov. Spravidla sa identifikujú a snažia sa dodržiavať pravidlá, ktoré im web nastaví.

Neškodní, ale príliš aktívni boti

Sem môžu patriť rôzne SEO, marketingové alebo analytické nástroje. Nemusia robiť nič škodlivé. Problém vznikne vo chvíli, keď začnú web prechádzať príliš intenzívne. Je to trochu ako návšteva, ktorá je síce vítaná, ale každých päť sekúnd zazvoní pri dverách.

Škodliví boti

Ich cieľom môže byť napríklad automatizované skúšanie prihlasovacích údajov, scraping obsahu, hľadanie zraniteľností alebo iné zneužitie webu.

Pre bežného majiteľa webu z toho vyplýva jednoduché pravidlo: **neblokujte návštevníka iba podľa mena, ak si nie ste istí, kto za ním skutočne je.**

A čo AI crawlery? Tie sú dobré alebo zlé?

Tu začína byť situácia zaujímavejšia. AI crawlery prechádzajú weby a zbierajú obsah pre AI služby. Ani medzi nimi však neplatí, že každý robí to isté. A to je dôležité najmä vtedy, ak chcete byť viditeľní nielen v Googli, ale aj v odpovediach AI vyhľadávačov.

Pekným príkladom je OpenAI. Aktuálne rozlišuje okrem iného:

- **GPTBot**, ktorý môže získavať obsah používaný pri tréningu generatívnych modelov,
- **OAI-SearchBot**, ktorý slúži na objavovanie stránok pre výsledky vyhľadávania v ChatGPT,
- **ChatGPT-User**, ktorý môže navštíviť stránku ako reakciu na konkrétnu požiadavku používateľa a nejde o klasický automatický crawler webu.

Nemusíte povoliť alebo zakázať všetko naraz. GPTBot môžete zablokovať pre tréning a OAI-SearchBotu nechať prístup, ak chcete zostať viditeľní v ChatGPT Search.

Otázka preto už neznie iba: „**Mám zablokovať AI botov?**“ Lepšia otázka je: „**Na čo tento konkrétny bot používa môj obsah a prináša mi to hodnotu?**“

Ak máte blog, magazín alebo e-shop, môže vám dávať zmysel, aby váš obsah AI vyhľadávače našli, citovali a priviedli k vám návštevníkov. To však neznamená, že ho musíte zároveň sprístupniť na tréning AI modelov.

OpenAI preto odporúča neblokovať OAI-SearchBot, ak chcete zvýšiť šancu, že sa váš obsah zobrazí v ChatGPT Search. Návštevy z ChatGPT navyše môžete sledovať v analytike cez parameter `utm_source=chatgpt.com`.

Pri AI crawleroch teda nejde len o technické nastavenie. Je to aj rozhodnutie o tom, kde a ako chcete mať svoj obsah viditeľný.

Ako môže bot spomaliť web, keď na vás nikto neútočí?

Na spomalenie webu nepotrebujete žiadny dramatický kybernetický útok. Stačí veľa požiadaviek.

Predstavte si svoj web ako obchod s piatimi pokladňami. Kým sú voľné, všetko funguje rýchlo. Keď sa všetky obsadia, ďalší zákazníci musia čakať.

Ak do predajne vojde skupina návštevníkov, ktorí chcú vidieť každý regál, každý produkt a každú možnú kombináciu, síce nič nenakúpia, no zamestnajú celý obchod. Podobne môže fungovať veľmi aktívny crawler. Na jednoduchom webe si ho možno ani nevšimnete. E-shop s tisíckami produktov, vyhľadávaním a filtrami je však iný príbeh.

Bot môže skúšať nasledovné príkazy:

?farba=modra

?farba=modra&velkost=m

... a následne stovky či tisíce ďalších kombinácií. Každá z nich znamená ďalšiu požiadavku na web. Ak ich príde priveľa, dostupná kapacita sa začne míňať a reálni zákazníci čakajú. Nikto pritom nemusí mať v úmysle váš web „zhodiť“. Automatizácia je jednoducho rýchlejšia než človek.

Ako zistíte, že za spomalením sú boti?

Prvým podozrením býva zvláštny nesúlاد: **web je pod záťažou, ale návštevnosť ľudí tomu nezodpovedá.**

Pri Google Analytics 4 je to ešte prirodzenejšie, pretože GA4 automaticky odfiltruje návštevnosť známych botov a spiderov. Analytika preto nemusí ukazovať všetky automatizované požiadavky, ktoré reálne dorazili na váš server.

Keď chcete zistiť viac, pomôžu serverové access logy, teda záznamy o požiadavkách, ktoré na web prichádzajú.

Na čo sa zamerať:

- opakuje sa jeden crawler veľmi často?
- prichádza nezvyčajne veľa požiadaviek v krátkom čase?
- crawler neustále prechádza URL adresy s filrami a parametrami?
- opakujú sa rovnaké alebo veľmi podobné stránky?
- v čase spomalenia rastie počet chýb alebo server nestíha odpovedať?

Dôležité je vedieť, že **Google Analytics nie je jediným miestom, kde sa návštevnosť webu odohráva**. Ak ste zákazníkom Websupportu, štatistiky návštevnosti si viete pozrieť priamo vo WebAdmine tu:

Ako botov obmedziť bez toho, aby ste si odrezali Google a AI vyhľadávanie?

Najlepším riešením väčšinou nie je „zablokovať všetko“.

Cieľom je pustiť na web návštevníkov, ktorí vám dávajú zmysel, a ostatných usmerniť alebo zastaviť.

1. Robots.txt = slušná smerovka pre crawlery

robots.txt je malý textový súbor, ktorý crawlerom hovorí, ktoré časti webu môžu alebo nemajú navštevovať.

Je to jednoduchý a užitočný prvý krok. Nie je to však bezpečnostný múr. Pravidlá `robots.txt` nie sú formou autorizácie prístupu a crawlery iba žiadajú, aby ich rešpektovali. Ak chcete prístup skutočne zablokovať, potrebujete inú vrstvu ochrany.

Ak napríklad nechcete povoliť OpenAI crawler určený na tréning, ale chcete zostať dostupní vo vyhľadávaní ChatGPT, nastavenie môže vyzeráť takto:

```
User-agent: GPTBot Disallow: / User-agent: OAI-SearchBot Allow: / User-agent:
* Disallow: /vyhľadavanie/
```

GPTBot tak prístup nemá, OAI-SearchBot áno. Ostatným botom zároveň môžete zakázať napríklad interné výsledky vyhľadávania, ktoré môžu vytvárať veľké množstvo URL adries.

Pravidlá si však vždy prispôsobte svojmu webu. Jedno univerzálne `robots.txt` pre všetkých neexistuje.

2. Blokovanie na serveri = keď prosba nestačí

Ak crawler `robots.txt` nerešpektuje alebo vás stále výrazne zťažuje, môžete ho blokovať priamo na úrovni servera. Na serveroch, ktoré používajú Apache a podporujú `.htaccess`, sa dá napríklad odmietnuť konkrétny User-Agent:

```
RewriteEngine On
RewriteCond %{HTTP_USER_AGENT} (SemrushBot|AhrefsBot|MJ12bot) [NC]
RewriteRule .* - [F,L]
```

Tu však už platí dvojnásobne: **najprv zistite, koho blokujete.**

User-Agent sa dá napodobniť, preto si nesprávnym nastavením môžete omylom zablokovať aj užitočného crawlera. Ak si nie ste istí technickým nastavením, radšej ho nechajte na administrátora.

A ešte jedna vec: server musí požiadavku najprv prijať, až potom ju môže odmietnuť. Pri veľkom množstve automatizovanej návštevnosti je preto lepšie zastaviť ju ešte skôr.

3. CDN a ochrana pred serverom = SBS pred budovou

Služby ako Cloudflare stoja medzi návštevníkom a vaším serverom a časť požiadaviek vyhodnotia ešte skôr, než sa dostanú k webu. Predstavte si ich ako SBSkára pred budovou, takže recepcia vo vnútri už nemusí riešiť každého, kto skúsi vojsť.

Cloudflare ponúka napríklad **AI Crawl Control**, cez ktorý môžete sledovať aktivitu AI crawlerov, jednotlivých botov povoliť alebo zablokovať a overiť, či rešpektujú pravidlá `robots.txt`.

Ďalšou možnosťou je **rate limiting**, teda obmedzenie počtu požiadaviek za určitý čas. Pomáha tak brzdiť napríklad scrapingové boty a chrániť web pred zbytočným preťažením.

Správny limit však nie je univerzálny. Závisí od typu webu, bežnej návštevnosti aj správania návštevníkov. E-shop počas výpredaja potrebuje iné nastavenie než jednoduchá firemná stránka.

CDN má navyše ešte jednu výhodu: často používaný statický obsah môže načítavať z cache, takže pri každej návšteve nemusí zťažovať pôvodný server.

Čo teda spraviť, ak máte podozrenie na botov?

Začnite otázkou: **Kto ku mne chodí a čo na webe robí?** Ak ide o užitočného crawlera, ktorý vám prináša viditeľnosť, nechajte ho pracovať. Ak ide o crawler, ktorého potrebujete iba usmerniť, začnite `robots.txt`. Ak nechcete poskytovať obsah na určitý účel (napríklad na tréning AI), nastavte pravidlá pre konkrétného crawlera, nie automaticky pre všetkých.

A ak bot ignoruje pravidlá alebo spôsobuje reálne problémy s výkonom, prichádza na rad serverové blokovanie, rate limiting, WAF alebo ochrana pred samotným serverom.

Najlepším výsledkom totiž nie je web bez botov. **Je ním web, na ktorý sa bez problémov dostanú zákazníci, vyhľadávače a služby, ktoré tam chcete mať.** Aby vám pritom zbytoční boti nespomaľovali web pre skutočných návštevníkov.

Najčastejšie otázky o botoch a crawleroch (FAQ)

Čo je bot a čo je crawler?

Bot je program, ktorý automaticky vykonáva úlohy na internete. Crawler je typ bota, ktorý systematicky prechádza webové stránky a ich odkazy, napríklad kvôli vyhľadávaniu alebo spracovaniu obsahu.

Môžu boti spomaliť web aj bez útoku?

Áno. Každá požiadavka spotrebúva určitú kapacitu. Ak crawler vytvorí veľké množstvo požiadaviek naraz, môže spomaliť web aj bez toho, aby išlo o úmyselný útok.

Prečo botov nevidím v Google Analytics?

Google Analytics 4 automaticky filtruje návštevnosť známych botov a spiderov. Kompletný obraz automatizovaných požiadaviek preto lepšie ukážu serverové logy.

Mám zablokovať Googlebot?

Vo väčšine prípadov nie. Ak chcete, aby bol váš web dostupný vo vyhľadávaní Google, potrebujete umožniť Googlu jeho obsah prechádzať. Ak sa návštevník iba vydáva za Googlebot, najprv overte jeho pôvod.

Mám blokovať AI crawlery?

Nie automaticky. Najprv zistite ich účel. Crawler pre tréning modelov môžete chcieť blokovať, zatiaľ čo crawler pre AI vyhľadávanie môžete ponechať povolený, ak chcete, aby sa váš web mohol zobrazovať a citovať v AI odpovediach.

Zdroje

Thales / Imperva: *2026 Bad Bot Report – Bad Bots in the Agentic Age*.

OpenAI: dokumentácia k crawlerom GPTBot, OAI-SearchBot a ChatGPT-User.

OpenAI: odporúčania pre publisherov a zobrazovanie webov vo výsledkoch ChatGPT Search.

IETF: *RFC 9309 – Robots Exclusion Protocol*.

Google Search Central: dokumentácia ku Googlebotu a overovaniu jeho identity.

Google Analytics: automatické filtrovanie známych botov a spiderov v GA4.

Cloudflare: AI Crawl Control, rate limiting a cache dokumentácia.

<https://www.websupport.sk/blog/2026/08/boti-crawlers-ai-crawlers>