

Proč analyzovat logy

Proč analyzovat logy

Každý provozovatel webových stránek by se měl zajímat o statistiky návštěvnosti, protože jen tak může posoudit úspěšnost webu a zaslíbeně rozhodovat o jeho budoucnosti. Měření návštěvnosti lze provádět několika způsoby, dnes se zaměříme na analýzu log souborů webového serveru a ukážeme si, jak si stojí ve srovnání s některými jinými způsoby měření.

Log soubory, zkráceně logy, jsou textové soubory obsahující záznamy o činnosti nějaké konkrétní aplikace. V případě webových serverů jsou do logů ukládány veškeré požadavky, které byly na server vzneseny. Zpětnou analýzou těchto dat pak můžeme zjišťovat cenné informace o fungování sledovaného webu.

Základní srovnání

V tomto článku se zaměříme na srovnání analýzy logů s měřicími systémy typu Google Analytics či NAVRCHOLU.cz, které k měření využívají speciální měřicí kód vložený do webových stránek. Nastavení takových systémů vyžaduje zásah do zdrojového kódu stránek, zatímco zaznamenávání logů je funkcí webového serveru a žádné úpravy nevyžaduje.

Systémy založené na měřicím kódu nabízí pouze hotové statistiky s omezenou možností dalšího zpracování, statistiky jsou však přístupné automaticky bez nutnosti dalších zásahů. Naproti tomu logy obsahují „surová“ data, která je nejprve nutná zpracovat. Výhodou tohoto přístupu je možnost pojmout zpracování dle vlastního gusta a výstupy si patřičně přizpůsobit. Data v měřicím systému jsou zpracována na základě dřívějších nastavení a zpětně z nich nové údaje nevytáhnete, logy jsou v tomto ohledu flexibilnější.

Logy také bodují zaručenou úplností dat v nich zaznamenaných. Funkční webový server zapíše do souboru každý požadavek, na který odpovídá. Naproti tomu systémy založené na měřicím kódu spoléhají na správné zpracování kódu na straně návštěvníkova prohlížeče, navíc dochází k odesílání dat na server třetí strany a celý proces je tak mnohem náchylnější k nepřesnostem. Odesílání a uchovávání dat na serveru poskytovatele měření návštěvnosti může být argumentem proti použití měřicích systémů v případě, kdy jsou statistiky považovány za citlivá data.

V logu naopak nenajdeme některé bližší informace o technickém vybavení návštěvníka, v tomto ohledu nabízí měřicí kódy více možností a detailnější informace.

Co logy nabízí navíc

Měřicí kód bývá zpravidla založen na JavaScriptu, do statistik se tudíž dostanou pouze ti návštěvníci, jejichž prohlížeč JavaScript zpracovává. To vylučuje jednak malou část ze skutečných návštěvníků (uživatelé s vypnutým JavaScriptem nebo nestandardním prohlížečem), jednak z takových statistik nevyčtete nic o různých automatizovaných nástrojích, jako například indexovacích robotech vyhledávačů. Pokud tedy chcete zjistit, jak často zkoumají vyhledávače Vaši stránku, jsou logy v podstatě jedinou možností. Pokud naopak o statistiky robotů zájem nemáte a zajímají Vás pouze skuteční uživatelé, je nutné roboty při zpracovávání logů vyfiltrovat, aby statistiky nebyly jejich chováním zkresleny.

Zřejmě nejzajímavější aplikací analýzy logů je sledování uživatelského chování. S využitím vhodných nástrojů lze velmi názorně zjistit, jakým způsobem se návštěvníci po webu pohybují, jak postupují při hledání určité informace a jak se jim to daří. Obzvláště zajímavé pak jsou analýzy složitějších úkolů, při kterých uživatel musí projít více stránek - typicky jde o proces objednávání v internetovém obchodě nebo vyplnění formuláře rozděleného do vícera kroků. Výstupy z takových analýz přesně ukazují, v kterých místech procesu návštěvníci tápou nebo se dokonce zcela ztrácejí a z webu odchází, aniž by dosáhli požadovaného cíle. Takové informace mohou výrazně napomoci ke zlepšení webu a zvýšení jeho obchodní účinnosti. Částečně lze některé údaje tohoto typu vyčíst i ze statistik získaných měřicím kódem, výhodou analýzy logů je velká flexibilita a možnost přizpůsobit analýzu konkrétnímu webu a požadovaným údajům.

Jak již bylo řečeno, do logů jsou zaznamenávány všechny požadavky na webový server. Díky tomu lze logy využít i ke sledování prvků, u kterých měřicí kód nelze jednoduše využít, jako jsou různé soubory ke stažení nebo obrázky. V logu také najdeme požadavky chybné, například žádost o poskytnutí neexistující stránky. Tímto způsobem lze tedy odhalit neplatné odkazy, v rámci samotného webu i z externích zdrojů, podobně jako další aspekty negativně ovlivňující fungování webu.

Nástroje

Nástroje pro zpracování logů lze rozdělit do dvou skupin. Zástupci první z nich se zpravidla instalují opět na webový server, logy zpracovávají v nastaveném intervalu a výsledky zpřístupňují přes webové rozhraní. Patří sem například nástroje AwStats a Webalizer, oba volně šiřitelné.

Pokročilejší možnosti pak nabízí speciální programy zaměřené na analýzu logů. Jejich pořízení nebývá právě levné, pro jednorázovou analýzu tak můžete využít služeb poradenských společností, které potřebným vybavením disponují.

Nastavení logů

Zaznamenávání do logů lze nastavit mnoha způsoby a spořiví správci často vypnou zaznamenávání některých údajů, v extrémních případech vypnou logování vůbec. Přístup k logům může být obtížný také v případě, že web neprovozujete na vlastním serveru a využíváte webhostingových služeb. Pokud tedy o analýze logů uvažujete, ujistěte se, že k souborům máte přístup. Z technického hlediska by v tom neměl být zásadní problém, pokud se poskytovatel technického zázemí Vašeho webu bude poskytnutí logů bránit, raději se nad důvody zamyslete. Rozsah zaznamenávaných údajů by pak měl odpovídat standardu Extended Common Log Format (ECLF), konkrétní tvar logu se může lišit podle použitého webového serveru.

Jak logy využívat

V čem je tedy analýza logů lepší oproti ostatním možnostem sledování návštěvnosti? Spíše by bylo vhodné říci, že každá metoda má svá specifika a své výhody. Jak jsme si ukázali, analýza logů nabízí informace o využívání webových stránek, které z jiných zdrojů není možné zjistit. Rozhodně však nejde o způsob univerzální, mnohé parametry je naopak možné sledovat pouze za použití jiné metody měření. Analýzu logů je možné provádět jednorázově za delší období, větší množství dat poskytne statisticky významnější výsledky. I v případě, že měření návštěvnosti provádíte jinou metodou, můžete z logů vyčíst nové informace, vedoucí k podstatnému zlepšení Vašeho webu.

Martin Smrt