

Datenschutz von Lehrern

Beitrag von „Thamiel“ vom 6. April 2016 20:20

Zitat von Herr Rau

Also bei der Homepage meiner Schule, die ich administrierte, funktioniert es. Schon immer. Wie überall. Ich denke, es ist ein Gerücht, dass das ein Gerücht ist.

Gut, dann geb ich dir ein Beispiel aus meinem eigenen Fundus: Ein [LogPNG](#) von heute mittag aus der immer noch laufenden Analysesoftware gezogen, die ich mal vor 15 Jahren programmiert habe, als mich derlei Sachen noch beruflich interessiert haben. Die Analysesoftware läuft immer noch, ich bin direkt stolz auf mich. Ich habs nur mal temporär auf die Homepage meiner Schule hochgeladen.

Das Teil zeigt die Zugriffaktivität der letzten 60 Tage. Jede Zeile ist ein Tag. Die maschinellen Zugriffe (soweit sich die Bots selbst zu erkennen geben) belaufen sich in dem Zeitraum übrigens auf knapp 65% aller Zugriffe. Was du da an Zugriffsstrichen siehst auf der rechten Seite sind alles Zugriffe, die sich *nicht* als maschinellen Ursprungs zu erkennen gegeben haben, sondern die sich als Mozilla, Chrome oder irgendwelche andere human gesteuerte Clientsoftware identifizierten.

Selbst beim schnellen überfliegen fallen einem dann aber die Ungereimtheiten auf:

1) Der Zugriffsspitzenschwerpunkt ist über dem Beobachtungszeitraum zwischen 0:00 und 5:00 morgens (oben von mir blau umkreist). Da sitzen überaus viele Menschen an ihren Rechnern!



2) Überall wo schwarze Blöcke in den Tageszeilen zu sehen sind, sind immens viele Zugriffe in kurzen Zeiträumen vorgenommen worden. Wenn man tiefer in die Logs einsteigt sind das Zugriffe im Sekundentakt. So schnell klickt kein Mensch auf eine Maus. Aber manchmal verraten sich diese Crawler durch ihre Muster, die sie hinterlassen. Beispielsweise die drei Kreise, die ich ab 21.3. dreimalig im 10-Tagesrhythmus um ca. 19:00 Uhr markiert habe. Übrigens wieder das Zugriffsmaximum während der Tageszeit und der Zugriffsclient soll Chrome gewesen sein. Wers glaubt...

Crawler haben ein Interesse, nicht als solche wahrgenommen zu werden. Um ihren Job zu machen brauchen sie nämlich die Website, auf die sie angesetzt werden, so angezeigt, wie ein menschlicher User sie angezeigt bekommt und nicht speziell für maschinelle Crawler mit zig Keywords gespickt, aber 0 Inhalt "maßgeschneidert".

Edit: Link korrigiert.