

SimilarWeb の数値は本当に信用できないか？

2018年10月15日
株式会社カドカワ
代表取締役社長 川上量生

第8回の検討会議において、海賊版サイトによる被害額算定の根拠とされる SimilarWeb を使ったアクセス数が信用できないという意見書が提出された。

意見書の中で SimilarWeb を否定する根拠として紹介されている英文サイトへのリンクの計4箇所を確認したところ、4箇所すべてにおいて指摘された内容と、大幅に異なる記述がみられた。

引用された英文資料と、それによる指摘内容をまとめると以下のとおりになる。

(1) SimilarWeb は米国の政府機関である FTC が実名入りで注意喚起の文書を出している（ぐららの悪質なサイトである）。

https://www.ftc.gov/system/files/documents/public_comments/2017/11/00043-141901.pdf

(2) SimilarWeb はハーバードビジネススクールの Edelman 教授が問題視して、利用すること自体に重大な責任があると指摘している。

<https://www.adweek.com/digital/six-companies-fueling-online-ad-crisis-150160/>

(3) SimilarWeb はデータを4割増しにしている。

<https://www.screamingfrog.co.uk/how-accurate-are-website-traffic-estimators/>

(4) Bloomberg の記事でも bot によるアクセス数の水増し問題の記事で SimilarWeb が言及されている。

<https://www.bloomberg.com/features/2015-click-fraud/>

以下、実際には元の英文サイトでどのように書かれているかを記す。

(1) **FTC が注意喚起** → (実際には) **FTC 主催のカンファレンスで学生が発表した資料**
立石委員提出資料 (第8回提出資料) の6頁目

FTC は、SimilarWeb を事例研究として実名入りで注意喚起

https://www.ftc.gov/system/files/documents/public_comments/2017/11/00043-141901.pdf

・ 上記 URL の PDF 文書は FTC 主催のカンファレンス PrivacyCON 2018 でノースイースタン大学の学生が発表した資料であり、**FTC が発表した資料ではない**。

(参照) <https://www.ftc.gov/news-events/events-calendar/2018/02/privacycon-2018>

(2) **ハーバードビジネススクール教授が問題視** → (実際には) **SimilarWeb ではなく別の会社**
資料6頁目に以下の記載

ハーバードビジネススクール Edelman 教授が問題視

SimilarWeb 社を利用すること自体に重大な責任があると指摘

・ 上記の記述の根拠と思われる記事は立石委員資料の8頁目に紹介されている以下の記事

<https://www.adweek.com/digital/six-companies-fueling-online-ad-crisis-150160/>

だが、**上記のような記述は存在しない**。

資料8頁目に以下の記載

これらの広告トラフィック販売会社のデータを元に分析を行っている3大会社が

・ SimilarWeb, eZanga, BlueLink

・ Harvard Business School professor Ben Edelman によれば、Smartmomstyle などこの会社を利用すること自体に重大な責任があると指摘

- ・ 原文を読むと前段は誤訳。

これらの広告トラフィック販売会社のデータを元に分析を行っている 3 大会社が SimilarWeb, eZanga, BlueLink

→ **SimilarWeb の数値によると**、eZanga と BlueLink が、（ボットを使っていると例示されているいくつかのサイト）の最大の参照元である。

- ・ 後段は原文でも Smartmomstyle（のようなサイト）と書いてあるが、文脈からは SimilarWeb は対象となっていない。

(3) **SimilarWeb は、データを 4 割増しに** → （実際には）**SimilarWeb の平均誤差は1%**
資料 8 頁目に以下の記載

<https://www.screamingfrog.co.uk/how-accurate-are-website-traffic-estimators/>

- ・ SimilarWeb は、データを 4 割増しに

・ 引用されている記事は、SimilarWeb、Ahrefs, SEMrush の 3 サイトの中で正確さ（Google Analytics のデータにもっとも近い）を比較した検証結果を載せている。

- ・ **もっとも正確なのは SimilarWeb** と記載されている。英国の 25 サイトを調査した結果、平均誤差は 1 % とある。

(4) **Bloomberg の Ad Fraud に関するレポートで、SimilarWeb についても言及** → （実際には）**たしかに言及はしているが・・・**

資料 8 頁目に以下の記載

<https://www.bloomberg.com/features/2015-click-fraud/>

- ・ Bloomberg の上記サイトには Ad Fraud に関する詳細なレポートがあり、やはり SimilarWeb についても言及

・ SimilarWeb を使って分析したと書いてあるが、SimilarWeb 自体に否定的な記述は見られない。以下、原文該当箇所。

but an analysis by SimilarWeb, a traffic-analysis firm, shows most of it arrived from a handful of identical-looking sites with names like Omnaling.com and Connect5364.com, each describing itself as “an advertising network technology domain.”

(まとめ)

引用された政府機関レポートとニュース記事である 4 つの英文資料すべてに根本的な間違いがあり、1 つは**そもそも政府機関のレポートではなく**、残り 3 つでは、むしろ **SimilarWeb の数値の信頼性を裏付ける内容**が書いてある。(2)の ADWeek と(4)の Bloomberg は、**SimilarWeb を利用して**記事を書いている。また、(3)では、**もっとも正確なトラフィック分析サイト**と結論づけられていることが分かる。

また他に引用されている日本語の個人ブログ記事をみても、**概ね正確な解析結果が出るとの評価も多い**と書いているものもある。

<https://nelog.jp/similarweb-data-is-from>

また、サイトによっては実際のアクセス数と大幅に乖離する場合があるという指摘もあるが、そのようなサイトはアプリからアクセスが多いサイトであるとの指摘もあり、今回の議論の対象となっている海賊版サイトは該当せず、また、その場合はアクセス数が過大ではなく過小になるわけであるから、いずれにせよ当てはまらない。

<http://netgeek.biz/archives/74929>

このことから報告書における被害金額、アクセス数の算出方法は妥当なものであると考える。数値の算出方法に過大評価される要因があることは否定しないが、それは漫画村の件に限らず、世の中で一般的に使われている同じような算出方法で得られた数値すべてに対していえることであり、今回の件だけとりあげて、おかしいと非難する理由はない。

また、そもそも漫画村の被害が甚大であったこと、マンガ文化が日本にとって大切かどうかということが、数値の誤差の有無で決定されるものではない。

以上