

Wpływ komunikatorów na wyniki oglądalności witryn internetowych – na podst. badania Megapanel PBI/Gemius.

Ze względu na odmienną charakterystykę użytkowania komunikatorów i zwykłych witryn, zdecydowano się na zmianę sposobu wyznaczania oglądalności witryn internetowych i komunikatorów. Niniejszy dokument ma na celu przybliżenie motywacji dla wprowadzonych zmian w metodyce wyznaczania oglądalności i zawiera przykładowe dane wynikowe przeprowadzonych analiz.

1. Wstęp

W niniejszym dokumencie jako komunikatory rozumiane są następujące aplikacje internetowe: Stefan, Tlen, Spik, AQQ oraz Gadu-Gadu. W dalszym opisie przedstawiono przykładowe dane wynikowe przeprowadzonych analiz.

Należy również mieć na uwadze, że ilekroć w niniejszym opisie będzie mowa o węzłach drzewa badania Megapanel PBI/Gemius, będzie to oznaczać robocze węzły wyliczeniowe widoczne tylko na etapie przygotowywania danych a nie stricte węzły, jakie użytkownicy badania widzą bezpośrednio w interfejsie badania.

Ponieważ wyłącznym celem prezentowanego opisu jest przedstawienie motywacji prowadzącej do zmiany metodyki badania, w dalszej części zdecydowano się na modelowy opis uzyskiwanych wyników (tj. zaniechano posługiwania się nazwami handlowymi). Witryny oznaczane są wielką literą, komunikatory – wielką literą z indeksem „kom”; jeśli w dotychczasowym badaniu węzeł bezpośrednio nadrzędny witryny był jednocześnie węzłem bezpośrednio nadrzędnym komunikatora, to witryna i komunikator zostały oznaczone tą samą literą (komunikator odpowiednio z indeksem „kom”), przykład:

- witryna porozmawiaj.pl i komunikator Porozmawiaj mają wspólny węzeł bezpośrednio nadrzędny. Zgodnie z przyjętą notacją witryna porozmawiaj.pl zostałaby oznaczona jako „P”, zaś komunikator Porozmawiaj – jako „P_{kom}”.

2. Wpływ komunikatorów na wyniki oglądalności witryn internetowych

Metoda estymacji rzeczywistej liczby użytkowników witryn internetowych w zadanym okresie czasu ([metoda estymacji](#)) odwołuje się do wskaźnika

tzw. dobrych cookie, tj. takich cookie, co do których mamy pewność, że istniały przez cały badany okres (cookie, które na pewno istniało przez cały badany okres to takie cookie, które na pewno istniało przed badanym okresem i po badanym okresie).

Jednym z czynników wpływających na poprawność wyników wykorzystywanego algorytmu estymacji jest zjawisko współkorzystania z różnych przeglądarek. Osoba, która korzysta z różnych przeglądarek, w danych site-centric jest reprezentowana przez różne cookie. Może to prowadzić do efektu tzw. wyłącznych cookie, tj. takich cookie, które w wybranym okresie badania w badaniu site-centric zostały zarejestrowane jedynie na wybranej witrynie (nie zostały zaś zarejestrowane na żadnym innym elemencie drzewa badania Megapanel PBI/Gemius). Dla danej witryny jej wyłączne cookie, co do których mamy pewność, że istniały przez cały badany okres stanowią zbiór wyłącznych dobrych cookie tej witryny.

Przyjmuje się, że opisany charakter korzystania przez internautów z tylko jednej witryny nie jest prawdziwym zjawiskiem. Taki efekt może wynikać z codziennego korzystania z jednej przeglądarki oraz przypadkowego uruchamiania innej (wówczas cookie przypadkowo uruchamianej przeglądarki może zostać zaobserwowane jedynie na jednej witrynie).

W przypadku witryn internetowych (tj. elementów drzewa badania Megapanel PBI/Gemius nie zawierających węzłów komunikatorów), stosunek liczby wyłącznych dobrych cookie danej witryny (UGC_w), do liczby wszystkich dobrych cookie zarejestrowanych na danej witrynie (GC_w) jest rzędu kilku procent. W Tabeli 2 przedstawiono wartości wielkości UGC_w/GC_w dla kilku największych witryn.

Tabela 2 Stosunek UGC_w/GC_w liczby wyłącznych dobrych cookie danej witryny (UGC_w), do liczby wszystkich dobrych cookie zarejestrowanych na danej witrynie (GC_w) dla wybranych witryn¹.

Witryna	UGC_w / GC_w
Witryna A	1.98%
Witryna B	1.55%
Witryna C	0.43%
Witryna D	0.83%
Witryna E	1.56%

W związku z rosnącym udziałem w rynku przeglądarek konkurencyjnych względem MSIE obserwuje się wzrost zjawiska wyłącznych dobrych cookie na komunikatorach (w badaniu site-centric komunikator może być reprezentowany przez cookie przeglądarki MSIE, która nie jest codziennie wykorzystywaną przeglądarką, więc jej cookie nie jest rejestrowane na innych witrynach). W wyniku kontroli danych źródłowych audytu site-centric, stwierdzono że w przypadku komunikatorów (za wyjątkiem komunikatora D_{kom}) obserwuje się znacznie większy udział ich wyłącznych dobrych cookie niż ma to miejsce w przypadku witryn. Analogicznie jak dla witryn, wyznaczono stosunek liczby wyłącznych dobrych cookie danego komunikatora (UGC_{kom}), do liczby wszystkich dobrych cookie zarejestrowanych na danym komunikatorze (GC_{kom}). W Tabeli 3 przedstawiono otrzymane wartości wielkości UGC_{kom}/GC_{kom} .

Tabela 3 Stosunek UGC_{kom}/GC_{kom} liczby wyłącznych dobrych cookie danego komunikatora (UGC_{kom}), do liczby wszystkich dobrych cookie zarejestrowanych na danym komunikatorze (GC_{kom})².

Komunikator	UGC_{kom} / GC_{kom}
Komunikator D_{kom}	0.00%
Komunikator G_{kom}	56.02%
Komunikator B_{kom}	18.87%
Komunikator H_{kom}	95.82%
Komunikator I_{kom}	23.41%

Drugą rozbieżność w charakterystyce użytkownika witryn i komunikatorów zaobserwowano badając liczbę odsłon przypadających na jednego użytkownika (użytkownika cookie).

Niech wielkość GP_v określa liczbę odsłon wykonanych przez dobre cookie (GC) w wybranym miesiącu na wybranym węźle obliczeniowym drzewa badania

Megapanel PBI/Gemius. Badając ruch na komunikatorach i węzłach dla nich nadrzędnych zaobserwowano, że dla niektórych komunikatorów stosunek GP_v/GC , a więc liczba odsłon wykonanych na komunikatorach przez dobre cookie zaobserwowane na tych komunikatorach przypadająca na jednego takiego użytkownika posługującego się dobrym cookie, jest znacznie większy niż analogiczna wielkość wyznaczona dla ich węzłów nadrzędnych. Wartości wielkości GP_v/GC dla komunikatorów i węzłów im nadrzędnych przedstawiono w Tabeli 4.

Tabela 4 Zestawienie liczby odsłon, wykonanych przez dobre cookie, przypadającej na jedno dobre cookie dla komunikatorów i węzłów im nadrzędnych. Wielkość GP_v/GC została wyznaczona bez pominięcia ruchu wyłącznych dobrych cookie komunikatorów. Pominięcie ruchu wyłącznych dobrych cookie komunikatorów prowadzi do wyznaczenia wielkości $(GP_v - GP_{v_{kom}})/(GC - GC_{kom})$.

Element w drzewie badania Megapanel PBI/Gemius	$\frac{GP_v}{GC}$	$\frac{(GP_v - GP_{v_{kom}})}{(GC - GC_{kom})}$
Węzeł nadrzędny komunikatora G_{kom}	207.37	119.42
Komunikator G_{kom}	1150.97	290.68
Węzeł nadrzędny komunikatora B_{kom}	184.10	184.11
Komunikator B_{kom}	139.09	132.47
Węzeł nadrzędny komunikatora H_{kom}	63.30	51.07
Komunikator H_{kom}	1367.26	720.64
Węzeł nadrzędny komunikatora I_{kom}	217.14	225.14
Komunikator I_{kom}	271.31	308.63
Węzeł nadrzędny komunikatora D_{kom}	110.89	110.89
Komunikator D_{kom}	1601.74	1601.74

Następnie dla każdego komunikatora wyznaczono jego wyłączne dobre cookie, tj. takie dobre cookie w danym miesiącu, które zaobserwowano jedynie na wybranym komunikatorze – ich liczbę określa wielkość GC_{kom} . Dla każdego komunikatora obliczono także liczbę odsłon na wybranym komunikatorze wykonanych przez jego wyłączne dobre cookie (GC_{kom}).

Dla komunikatora H_{kom} wyłączne dobre cookie stanowią 95.8% wszystkich dobrych cookie tego węzła, ale już tylko blisko 0.91% dobrych cookie węzła wyliczeniowego nadrzędnego dla komunikatora H_{kom} . Odsłony wykonane przez wyłączne dobre cookie komunikatora H_{kom} stanowią 97.8% wszystkich odsłon wykonanych na tym komunikatorze przez dobre cookie (w tym cookie, które zaobserwowano na innych węzłach); tymczasem odsłony wykonane przez wyłączne dobre cookie komunikatora H_{kom} stanowią już 20% wszystkich odsłon wykonanych na jemu nadrzędnym

¹ Na podstawie danych badania site-centric realizowanego przez firmę Gemius w okresie listopad 2008 - marzec 2009.

² Na podstawie danych badania site-centric realizowanego przez firmę Gemius w okresie marzec – maj 2009.

węźle wyliczeniowym przez dobre cookie tego węzła. Czyli dla węzła nadrzędnego komunikatora H_{kom} pominięcie ruchu wyłącznych dobrych cookie komunikatora H_{kom} powoduje zmianę liczby dobrych cookie o 0.91p.p. oraz zmianę liczby dobrych odsłon o 20p.p.

Dla komunikatora G_{kom} wyłączne dobre cookie stanowią blisko 56% wszystkich dobrych cookie tego komunikatora, ale już tylko blisko 5.2% dobrych cookie nadrzędnego węzła wyliczeniowego komunikatora G_{kom} . Odsłony wykonane przez wyłączne dobre cookie komunikatora G_{kom} stanowią 88.9% wszystkich odsłon wykonanych na tym komunikatorze przez dobre cookie (w tym cookie, które zaobserwowano na innych węzłach); tymczasem odsłony wykonane przez wyłączne dobre cookie komunikatora G_{kom} stanowią już 45.38% wszystkich odsłon wykonanych na jemu nadrzędnym węźle wyliczeniowym przez dobre cookie. Czyli dla nadrzędnego węzła wyliczeniowego komunikatora G_{kom} pominięcie ruchu wyłącznych dobrych cookie komunikatora G_{kom} powoduje zmianę liczby dobrych cookie o 5.2p.p. oraz zmianę liczby dobrych odsłon o 88.9p.p.

Pominięcie ruchu wyłącznych cookie komunikatorów przy wyznaczaniu liczby odsłon, wykonanych na węzłach przez dobre cookie zaobserwowane na tych węzłach, przypadających na jedno takie dobre cookie prowadzi do wyznaczenia wielkości $(GP_v - GP_{v_{kom}})/(GC - GC_{kom})$. Otrzymane wyniki zestawiono w Tabeli 4. Pominięcie ruchu wyłącznych cookie komunikatorów (pominięcie dobrych cookie, które zaobserwowano jedynie na komunikatorze oraz odsłon wykonanych przez takie dobre cookie) powoduje, że zmienia się liczba odsłon przypadających na użytkownika (dokładniej mowa o liczbie „dobrych” odsłon przypadających na dobre cookie). Zatem dobre cookie zaobserwowane jedynie na komunikatorach wykonują na komunikatorze odsłony dużo częściej niż dobre cookie, które zaobserwowano także na innych węzłach. Świadczy to o tym, że ze względu na odmienną charakterystykę użytkownika, na komunikatorach odsłony są generowane częściej niż na węzłach nie zawierających komunikatorów. Potwierdzają to wyniki badania gemiusTraffic – dla konta komunikatora H_{kom} czas trwania odsłony wynosi ok. 5 min, zaś dla najpowszechniejszego skryptu komunikatora G_{kom} czas trwania odsłony wynosi ok. 3 min i są to wielkości praktycznie niezmiennie. Możliwym wyjaśnieniem jest to, że odsłony na tych komunikatorach są generowane automatycznie (bez ingerencji użytkownika).

Odmienna charakterystyka użytkownika komunikatorów i zwykłych witryn, motywuje do opracowania nowej metody wyznaczania zasięgów na węzłach badania Megapanel PBI/Gemius.

3. Nowa metoda wyznaczania oglądalności

Nowa metoda wyznaczania oglądalności witryny zakłada, że na jej wartość nie wpływa ruch z komunikatorów (ruch z komunikatorów jest pomijany przy wyznaczaniu oglądalności witryny).

Dla wszystkich audytowanych węzłów nie zawierających komunikatorów (tj. takich węzłów, które nie były nadrzędnymi względem komunikatorów) wzrost oglądalności wynosi 16.26% w stosunku do wartości wyznaczanych dotychczasową metodą³. Jest to efekt związany ze zmianą estymowanej liczby użytkowników internetu.

Dla audytowanych węzłów, które dotychczas zawierały komunikatory (węzłów nadrzędnych dla komunikatorów), pominięcie ruchu komunikatorów prowadzi do zmiany oglądalności także o ok. 16% w stosunku do wartości wyznaczanych z dotychczasową metodą³.

Zasięg komunikatorów jest wyznaczany w odniesieniu do estymowanej liczby użytkowników internetu, która jest zgodna z estymowaną liczbą użytkowników witryn internetowych. Prowadzi to do wzrostu zasięgu o 16.26% w stosunku do wartości wykorzystywanych dotychczasową metodą³.

W Tabeli 5 zamieszczono porównanie zasięgów wyznaczonych zgodnie z dotychczasową i nowo przyjętą metodą dla wybranych węzłów wyliczeniowych badania Megapanel PBI/Gemius.

Tabela 5 Zestawienie wartości zasięgów dla wybranych węzłów. Indeks „kom” określa wielkości wyznaczone zgodnie z dotychczasową metodą uwzględniająca ruch na komunikatorach; wielkości bez indeksu „kom” zostały wyznaczone z pominięciem ruchu komunikatorów³.

Witryna	Z_{kom}	Z
Witryna A	61.18%	71.13%
Witryna B	54.19%	63.01%
Witryna C	47.36%	55.06%
Witryna D	48.03%	55.84%

³ Na podstawie danych badania site-centric realizowanego przez firmę Gemius w okresie styczeń - marzec 2009.