

Opis JAR Flex

01.07.2025

1. JAR Flex – wprowadzenie

Wdrożenie JAR Flex ma na celu wzmocnienie jakości badania w związku ze wzrastającym udziałem odsłon z przeglądarek nieobsługujących Third-Party Cookies (TPC).

Przez wiele lat lobbowano o zwiększenie poziomu prywatności w sieci, w wyniku czego niektóre przeglądarki zrezygnowały z obsługi plików Third-Party Cookie. Obecnie nie są one już domyślnie obsługiwane przez przeglądarki: Safari, Samsung Browser, Microsoft Edge i Mozilla Firefox, które odpowiadają za ok. 25% wszystkich odsłon z przeglądarek.

Dodatkowo wskazać trzeba na zjawisko powolnego słabnięcia pozycji przeglądarki Chrome i wzrostu korzystania przez użytkowników z innych narzędzi. To zmienia charakterystyki i schematy korzystania z Internetu. Chcąc zachować najwyższą jakość badania i wiarygodność danych musimy – na ile to możliwe - odejść od traktowania użytkowników przeglądarki Chrome jako benchmarku statystycznego i włączyć do puli badawczej użytkowników innych przeglądarek. Bez odejścia od technologii korzystającej wyłącznie na plikach TPC i bazowaniu na przeglądarce Chrome byłoby to niemożliwe.

JAR Flex to rozwiązanie, które zapewnia wzrost jakości badania, lepsze odwzorowanie charakterystyki ruchu internetowego i korzystania z Internetu przez użytkowników oraz gwarantuje ciągłość badania, bez względu na to, jaka część uzyskiwanych odsłon będzie pochodziła z przeglądarek nieobsługujących Third-Party Cookies. Dzięki temu nasze badanie staje się bardziej niezależne od zmieniającej się technologii, jednocześnie zachowując wysoką jakość wyników.

2. Opis metody

2.1 Jak działa JAR Flex ?

1) Wyniki dla odsłon, czasu i wizyt obliczane są przy wykorzystaniu plików First-Party Cookies (FPC) ze wszystkich przeglądarek.

2) Wskaźnik Real Users obliczany jest na dwa sposoby:

- przy użyciu tradycyjnej metody Estimated Cookies (EC), opartej na identyfikatorach TPC (tzw. identyfikator BID) w przypadku przeglądarek, które nadal je obsługują. Zmiana obejmuje jednak przejście z dotychczas stosowanej wersji tej metody - EC_global - na wersję EC_standard.
- przy użyciu metody Browsers Number (BN) w przypadku pozostałych przeglądarek.

3) Metoda Cookie Matching (CM) wykorzystuje FPC poszczególnych przeglądarek. W przypadku tych z nich, które nadal obsługują identyfikator BID, zachowane zostaną rzeczywiste połączenia między FPC wskazanymi przez wspólny identyfikator BID.

2.2 Metody w JAR Flex

JAR Flex składa się z dwóch Metod:

- 1) Metody Cookie Matching (CM)
- 2) Metody Browsers Number (BN)

Cookie Matching

Cookie Matching" (CM) to metoda oceny prawdopodobieństwa, że pliki First-Party Cookies (FPC) pochodzące z różnych domen, które obserwujemy jako oddzielne, w rzeczywistości pochodzą z jednego urządzenia, a następnie łączenia ich w odpowiednie grupy.

Prawdopodobieństwo, że dwa FPC pochodzą z jednego urządzenia, jest obliczane przez klasyfikator uczenia maszynowego, który posiada poniższą zasadę działania:

- jeśli dwa identyfikatory są często obserwowane w tym samym miejscu (tj. pod tym samym adresem IP) w tym samym czasie, prawdopodobieństwo, że pochodzą z jednego urządzenia wzrasta,
- jeśli dwa identyfikatory są często obserwowane w różnych miejscach w tym samym czasie, prawdopodobieństwo, że pochodzą z jednego urządzenia maleje.

Na podstawie tego prawdopodobieństwa łączymy FPC w grupy zwane „communities”, które są odpowiednikami Third-Party Cookies. Próbujeśmy odtworzyć wszystkie TPC na danym rynku, jednak tym, co rzeczywiście leży w zakresie naszego zainteresowania, jest nasz Cookie Panel. Aby zaprezentować profil społeczno-demograficzny zachowujemy więc tylko te communities, które mają przydzielony wypełniony kwestionariusz naszego badania, a więc communities ze znaną demografią.

1) Segmentacja FPC

Wybór FPC opiera się na rozmiarze adresu IP (definiowanym jako liczba FPC z danego IP), z którego pochodzą. Jeśli zaobserwujemy, że w którejkolwiek z domen na jednym adresie IP jest zbyt wiele aktywnych FPC w badanym okresie, wówczas nie używamy ich do łączenia, ze względu na zbyt dużą niepewność klasyfikatora.

2) Tworzenie zestawu potencjalnych par

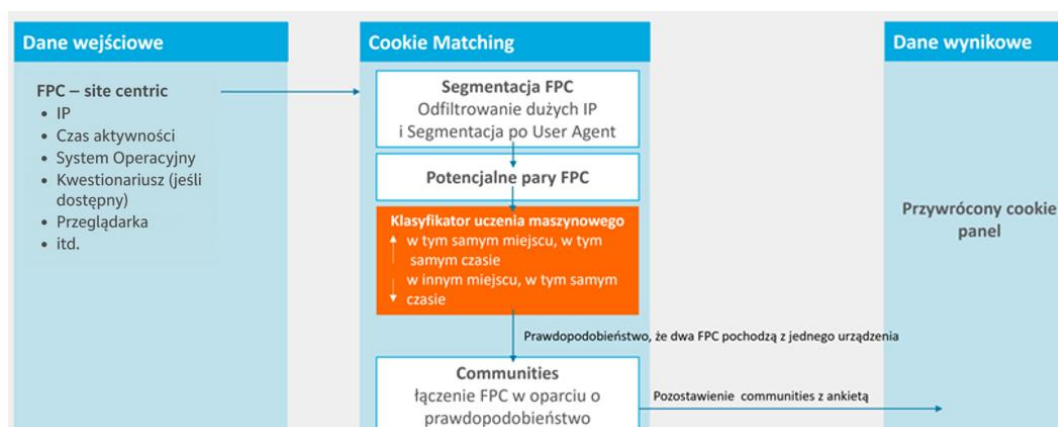
FPC wybrane w poprzednim kroku, aktywne w danym okresie, są oceniane przez klasyfikator łącznie tylko wówczas, gdy dzielą przynajmniej jeden wspólny adres IP. Nie mogą one również pochodzić z tej samej strony, w tej samej przeglądarce. Pary z analizowanych okresów są następnie łączone w jeden zestaw, bez duplikatów. Następnie obliczane są różne atrybuty dotyczące aktywności pod tymi samymi adresami IP.

3) Klasyfikator uczenia maszynowego

W tym modelu jako zestawu uczącego używamy danych z przeglądarek, które obsługują TPC (m.in Chrome). Dla dwóch zestawów FPC - pochodzących i niepochodzących z tego samego TPC - wybierany jest zestaw potencjalnych par, a predyktory są obliczane w taki sam sposób, jak w punktach 1 i 2. Model jest następnie uczony na tych danych. Na koniec, przy użyciu zestawu potencjalnych par, model przewiduje prawdopodobieństwo pochodzenia FPC z jednego urządzenia.

4) Algorytm łączenia par w „communities”

Łączenie FPC w communities odbywa się iteracyjnie. Dane wejściowe dla tego kroku to zestaw par FPC - FPC (gdzie każdy FPC może wystąpić wiele razy) i przypisane im prawdopodobieństwo. Jednak nie wszystkie pary mogą zostać połączone, a próg dopuszczalności ich łączenia (minimalne prawdopodobieństwo) wzrasta w każdej kolejnej iteracji. Na każdym etapie dbamy również o zachowanie warunku mówiącego, że w jednym community nie mogą znajdować się FPC istniejące w tej samej domenie w tym samym czasie. Niektóre z wynikowych communities będą zawierać FPC z przypisanym kwestionariuszem badania. Ta część stanowić będzie odtworzony cookie panel przeglądarek, które nie obsługują TPC. Jeśli do danego community trafi więcej niż jeden FPC z kwestionariuszem, to community nie będzie włączone do odtwarzanego cookie panelu.



Browsers Number

"Browsers Number" (BN) to zestaw metod obliczania na podstawie danych site-centric szacunkowej liczby przeglądarek (do tej pory nazywanych Estimated Cookies (EC)) dla domen, grup oraz węzła łączącego cały audytowany ruch internetowy, bez użycia plików Third-Party Cookies (TPC). BN wykorzystuje wyłącznie dane dostępne po wycofaniu TPC. Źródłami danych wykorzystywanymi przez BN zamiast TPC są: First-Party Cookies (FPC) i adresy IP.

Browsers Number jest liczone analogicznie, jak w metodzie EC Standard. Oznacza to, że przy wyliczaniu szacunkowej liczby przeglądarek stosunek wszystkich odsłon do odsłon pochodzących od identyfikatorów uznanych za „dobre” jest oparty na wartościach składowych dotyczących danego kanału mediowego (domeny, serwisu lub grupy). W dotychczas stosowanej w Badaniu metodzie EC Global stosunek ten był wyliczany nie na poziomie poszczególnych kanałów mediowych, ale na poziomie całego audytowanego ruchu łącznie.

Metoda EC Global wymagała reprezentatywnej próbki plików cookie, tak zwanych "dobrych plików cookie", których charakterystyka była ekstrapolowana na wszystkie pliki cookie (przy użyciu stosunku liczby dobrych cookie do liczby odsłon wykonanych przez te cookie). BN również potrzebuje reprezentatywnej grupy, której charakterystyka byłaby przekładana na cały ruch, ale tym razem tę rolę pełnią reprezentatywne grupy FPC i adresów IP. Znajdowanie takich reprezentatywnych podgrup jest sercem metody BN.

Metody działające w ramach BN są zaprojektowane i zoptymalizowane dla kilku wymiarów:

- dwóch dotyczących czasu - dnia i pseudomiesiąca (28 dni)

oraz

- dwóch dotyczących obiektu zainteresowania - pojedynczej domeny i wielu domen (w tym węzła całego audytowanego internetu).

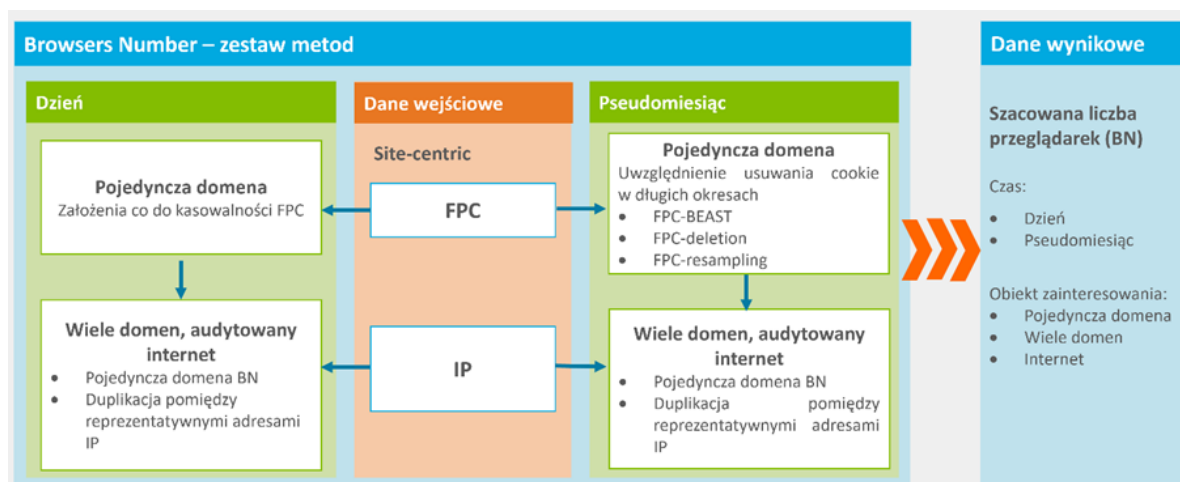
W sumie istnieją więc 4 kombinacje, a więc 4 nieco różne, ale dobrze uzasadnione sposoby podejścia do wyszukiwania reprezentatywnych grup danych:

1) Dzień, pojedyncza domena - używanie tylko plików FPC z pewnymi założeniami dotyczącymi ich usuwania.

2) Dzień, wiele domen, audytowany internet - z wykorzystaniem informacji o BN dla pojedynczych domen (tj. punkt nr 1) oraz duplikacji pomiędzy reprezentatywnymi adresami IP dla kilku największych domen.

3) Pseudomiesiąc, pojedyncza domena – stosowanie odpowiednika metody BEAST* w celu wyznaczenia reprezentatywnej próby FPC, aby obliczyć na jej podstawie liczbę przeglądarek.

4) Pseudomiesiąc, wiele domen, cały audytowany internet - wykorzystanie informacji o BN dla pojedynczych domen w dłuższych okresach (tj. punkt nr 3) oraz duplikacji pomiędzy reprezentatywnymi adresami IP największych domen.



* BEAST (Browser Estimation Algorithm Standard) - jest to algorytm predykcyjny, pozwalający ocenić czy dany plik cookie okaże się w przyszłości plikiem good cookie. Przewidywanie to dokonywane jest na podstawie poziomu wcześniejszej aktywności danego cookie oraz czasu, jaki upłynął od dnia, gdy widzieliśmy go po raz ostatni.

3. Wpływ na wyniki

Wdrożenie metodologii JAR Flex wpływa na wyniki w następujący sposób:

- przejście z metody EC Global na metodę EC Standard wpływa na obniżenie liczby użytkowników tam, gdzie stosunek wszystkich odsłon do dobrych odsłon (allPV/goodPV) jest niższy od średniej w całym internecie audytowanym oraz wpływa na ich podwyższenie tam, gdzie jest on wyższy od tej średniej.
- gdy udział odsłon z agregatora Google Discover jest wysoki, to domena traci liczbę użytkowników, a gdy udział jest niski, to domena ich zyskuje. Dzieje się tak, ponieważ kiedy domena czerpie dużo ruchu z Discover, który otwiera strony we wbudowanej w aplikację Google przeglądarkę Chrome, to wpływa to na niższy niż średni w internecie odsetek odsłon non-BID. Przejście na EC-standard i dodanie do puli badawczej ruchu z innych przeglądarek niż Chrome skutkuje tym, że domeny z dużym udziałem od Discovera wykazują spadki liczby użytkowników.
- kanały mediowe posiadające restrykcyjne podejście do polityki prywatności (na których domyślnym scenariuszem lub ze względu na dbałość o prywatność użytkownika jest odmowa zgody na używanie 3rd-party cookie) będą zyskiwać liczbę użytkowników, ponieważ udział odsłon non-BID będzie tam wyższy. Rozwiązanie JAR Flex w takim przypadku wyrównuje szanse wydawców, nie stawiając w mniej korzystnej pozycji tych, którzy w większym stopniu dbają o prywatność użytkowników.

Gemius S.A.
ul. Domaniewska 48
02-672 Warszawa, Polska

Kontakt:
mediapanel@gemius.com
+ 48 22 390 90 90
+ 48 22 378 30 50