

JAR Flex

K 1. září 2025 dochází v projektu měření návštěvnosti internetových stránek NetMonitor k zavedení metodiky označované jako „JAR Flex“. Jejím hlavním přínosem je rozšířená schopnost identifikace uživatelů o některé internetové prohlížeče, které v současné době již nepodporují identifikaci za pomoci tzv. cookies třetích stran. Metodika dále reaguje na některé problémy a výzvy měření návštěvnosti internetu. Česká republika není první ani jedinou zemí, kde k nasazení JAR Flex došlo. Do tohoto okamžiku byla tato metodika schválena a implementována v řadě zemí a projektů včetně Belgie, Dánska, Portugalska, Maďarska nebo Lotyšska a další trhy budou v průběhu letošního roku následovat. Hlavním důvodem implementace po několik let vyvíjené metodiky JAR (Joint Alternative Recognition) je příprava projektu NetMonitor na možné budoucí scénáře znemožnění nebo významného omezení schopnosti identifikace internetových uživatelů za pomoci aktuální technologie. Zároveň se tímto krokem rozšiřují možnosti pro rekrutaci panelů o tyto prohlížeče a i pro samotný výpočet základní metodiky Reálných Uživatelů (RU).

JAR Flex kombinuje použití identifikace uživatelů pomocí cookies třetích stran (3rd party cookies, TPC) a nově i cookies prvních stran (1st party cookies, FPC).

Metodické změny, které JAR Flex přináší:

1) Výpočet RU

JAR Flex kombinuje pro výpočet metriky Reálných uživatelů (RU) TPC např. z prohlížeče Chrome nebo Opera tak jak je tomu i v aktuální metodice, s FPC z dalších prohlížečů – např. Firefox, Edge. V případě ukončení nebo omezení TPC ve zmíněných prohlížečích bude pro výpočet RU používána identifikace založená na FPC. Tento přístup nás a naše klienty staví do pozice, ve které jsme připraveni na jakýkoli vývoj v otázkách dostupné identifikace uživatelů. JAR Flex omezuje výraznou závislost výpočtu RU na TPC v Google Chrome, rozšiřuje bázi prohlížečů pro tento výpočet, zavádí opět možnost rekrutovat uživatele do panelů v prohlížečích Firefox a Edge a pozitivně tak přispívá k jejich velikosti.

1a) EC Standard (pro TPC)

U TPC dochází k dílčí změně, která má ale významný dopad na data. Důležitým mezikrokem algoritmu kalkulace RU je výpočet tzv. Estimated Cookies (EC). Do tohoto výpočtu mimo jiné vstupuje poměr všech měřených Pageviews a tzv. „good“ Pageviews.

$$EC = GC * \frac{ALL\ PVs}{Good\ PVs}$$

Nově je tento poměr počítán pro každý uzel zvlášť a je tak pro něj unikátní. Tento přístup je označován jako EC Standard.

EC Standard je z principu odolnější vůči neočekávaným a významným změnám v měření, které ovlivňují celý měřený uzel Internet. Příkladem je např. vypadnutí celého velkého Provozovatele z měření, kdy dojde k ovlivnění poměru APV/GPV celého uzlu Internet a tím i k dopadu na další měřené subjekty.

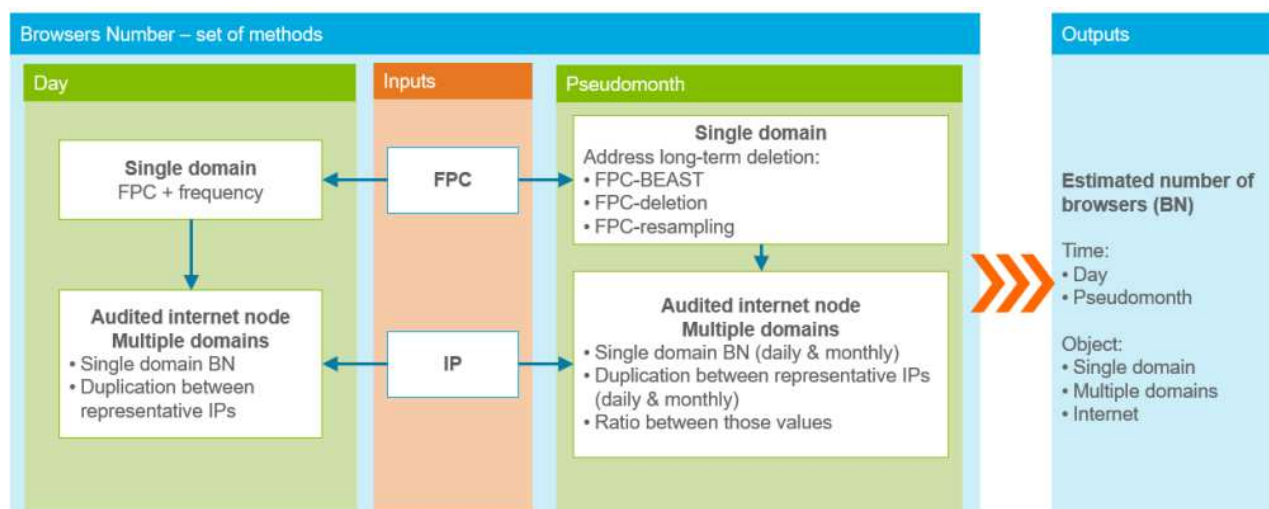
Druhým faktorem mluvícím pro EC Standard je možnost budoucích úprav lišt pro udělení souhlasu uživatelů se zápisem cookies do jejich prohlížečů. V případě, kdy jeden významný nebo i několik větších či menších subjektů změní (zprísnil, rozvolní) pravidla pro souhlas s ukládáním cookies, může opět dojít ke změně poměru APV/GPV a tím i k ovlivnění dat subjektů, které tato pravidla nezměnily.

1b) Browsers Number (pro FPC)

JAR Flex využívá pro výpočet RU tzv. Browsers Number (BN). V principu se jedná o stejnou věc, jako jsou EC – počet prohlížečů, které by systém identifikoval, pokud by neexistoval problém s mazáním a akceptací cookies. V případě nutnosti použít 1st party cookies (FPC) není nadále možné používat současný přístup založený na výpočtu Estimated Cookies, protože v případě FPC chybí informace o propojení jednotlivých cookies na různých měřených stránkách, ale mnohdy i mezi cookies na jedné měřené doméně v různých obdobích dnů či týdnů. Produkce hodnot BN v metodice JAR Flex pracuje s daty všech měřených FPC a informací o jejich aktivitě (pageviews), čímž se liší od současné metodiky EC, která sice rovněž pracuje se všemi měřenými pageviews, ale dále na úrovni cookies používá pouze jejich podkupinu (tzv. “good” cookies). Pokud zcela zjednodušíme princip výpočtu RU, pak lze říci, že každodenní produkce dat v systému gemiusAudience vyžaduje (mimo jiné) čtyři základní čísla:

- Hodnotu EC pro danou stránku za poslední den
- Hodnotu EC pro danou stránku za posledních 28 dní
- Hodnotu EC pro uzel Internet za poslední den
- Hodnotu EC pro uzel Internet za posledních 28 dní

Zatímco současná metodika je schopna tyto hodnoty spočítat jednoduchým postupem za použití jedné rovnice aplikovatelné na všechny uvedené případy, JAR limitovaný omezenými možnostmi FPC, které jsme nastínili výše, musí pro každý případ použít jiný přístup.



2) Cookie Matching

Metoda vyhodnocování pravděpodobnosti, že FPC pocházející z různých domén, které pozorujeme jako samostatné, ve skutečnosti pocházejí z jednoho zařízení, a následně je podle toho spojujeme do skupin. Pravděpodobnost dvou FPC pocházejících z jednoho zařízení vypočítává klasifikátor strojového učení, který má velmi jednoduchou myšlenku:

- o pokud jsou často vidět dva identifikátory na stejném místě (tj. pod stejnou IP), ve stejnou dobu, zvyšuje se pravděpodobnost, že pocházejí z jednoho zařízení,
- o a pokud jsou často vidět dva identifikátory na různých místech ve stejnou dobu, pak se pravděpodobnost, že pocházejí z jednoho zařízení, snižuje.

Na základě této pravděpodobnosti spojujeme FPC dohromady do skupin nazývaných Komunity, což jsou ekvivalenty TPC. Snažíme se obnovit všechny TPC na daném trhu, ale to, co nás ve skutečnosti zajímá, je Cookie Panel, takže si ponecháváme pouze komunity s vyplněným dotazníkem (tj. komunity se známou demografií). CM se neomezuje na spojování identifikátorů z prostoru jednoho identifikátoru. Dokáže tedy propojit weby, aplikace a webview in-app browsery pocházející z jednoho zařízení a jde tak ještě dále než jen k obnově TPC.

1. Segmentace FPC

Výběr FPC je založen na velikosti IP adresy, ze které pocházejí. Pokud zjistíme, že na některé z domén na jedné IP adrese je v analyzovaném časovém období (1d, 3h, 1h) příliš mnoho aktivních FPC, nepoužijeme je pro slučování z důvodu příliš vysoké nejistoty.

2. Vytvoření sady potenciálních dvojic

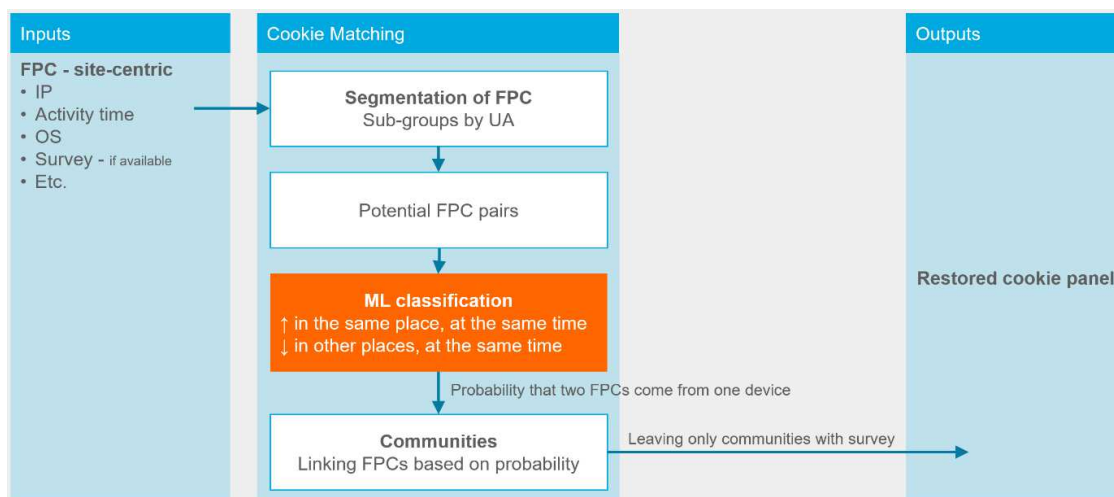
FPC vybrané v předchozím kroku, aktivní v daném časovém období, jsou klasifikátorem hodnoceny společně pouze pokud sdílely alespoň jednu IP adresu. Nemohou také pocházet ze stejného webu a ve stejném prohlížeči. Páry ze tří analyzovaných období se pak spojí do jediné sady bez duplikátů. Poté jsou vypočítány různé atributy týkající se aktivity pod stejnými IP adresami (seznam atributů je důvěrné know-how Gemiusu).

3. Klasifikace strojového učení

Pro dvě sady FPC pocházející a nepocházející ze stejné TPC se vybere sada potenciálních párů a prediktory se vypočítají stejným způsobem jako v bodech 1 a 2. Na těchto datech je pak model trénován. Nakonec pomocí sady potenciálních párů model předpovídá pravděpodobnosti, že FPC pocházejí z jednoho zařízení.

4. Algoritmus pro spojování párů v komunitě

Připojování FPC do komunity se provádí iterativně. Vstupními daty pro tento krok je sada párů FPC - FPC (kde každá FPC se může vyskytovat vícekrát) a přiřazená pravděpodobnost. Ne všechny páry se však smí sloučit, práh (minimální pravděpodobnost) se v každé iteraci zvyšuje. V každé fázi také dbáme na zachování podmínky, že v jedné komunitě nemohou existovat různé FPC na stejné doméně současně. Některé z výsledných komunit budou zahrnovat FPC s vyplněným dotazníkem. Tato část bude znovu vytvořeným cookie panelem prohlížečů, které nepodporují TPC. Pokud se do Komunity dostane více než jedna FPC s vyplněným dotazníkem, nebude tato komunita brána v úvahu.



3) Estimated Visits

Metoda dopočítávání Návštěv. Odhad návštěv pro non-cookie traffic na uzlu na základě jeho cookie trafficu (toto je důležité pro uzly, kde se úroveň RU blíží úrovni počtu Visits). Metodika Estimated Visits pro non-cookie část trafficu zvýší počet Visits a tím méně uzlů má tedy snížené RU kvůli korekci RU vs. Visits.

4) PRES

Pro stanovení univerza internetových uživatelů 10+ v ČR se zavádí metodika PRES (Population and Reach Estimated Smoothly) – která využívá data strukturálního výzkumu jako základ, ale navíc bere v potaz i další zdroje dat jako je např. site-centric audit, dotazníky a software panel - tak, aby reflektovala adekvátní trendy. Díky tomu zlepšujeme kvalitu výzkumu, jsme schopni odhadovat velikosti populací pro následující měsíce a provádět plynulejší změny mezi jednotlivými vlnami strukturálního výzkumu.

5) Platforma Mobilní zařízení

Bude zavedena platforma Mobilní zařízení místo samostatných telefonů a tabletů.

6) Věková rozmezí v procesu vážení

U nejmladší věkové kategorie se změní rozmezí (10-17, 18-24 místo nyníšších 10-14, 15-24). Týká se jen procesu vážení, v datových výstupech se nic nemění.