

Metodický dokument – Metodika výskumu AIMmonitor

Výskum AIMmonitor je postavený na riešení **gemiusAudience™**, ktoré využíva vlastnosti dát získaných pomocou site-centric spôsobov zberu informácií, spolu s oddelenými dátami štruktúrálnej štúdie (charakteristiky populácie).

Výskum poskytuje informácie o sociálnom a demografickom profile návštevníkov internetu (konzumentov obsahu a služieb) a umožňuje rozdelenie výsledkov na samostatné dátové sady pre každú platformu, napr. PC, smartfóny, tablety, smart TV. Výsledky **gemiusAudience™** umožňujú:

- odhadnúť skutočný počet internetových užívateľov (Reálnych užívateľov, ďalej ako RU) s využitím **BrowserID™** identifikátora a algoritmu **BEAST™**, ako aj reach akéhokoľvek individuálneho média zapojeného do štúdie
- poskytnúť sociálny a demografický profil internetových užívateľov na základe dát panelistov
- zahrnúť základné merané metriky ako je napr. zobrazenia stránky, čas a návštevy, ako aj akékoľvek z nich odvodené metriky
- vypočítať výsledky na zozbieraných dátach vysokej kvality vďaka procesu kontroly kvality a filtrovania dát
- poskytnúť porovnateľné dáta o správaní užívateľa bez ohľadu na prehliadač, systém a zariadenie
- agregáciu a porovnateľnosť metrík medzi jednotlivými účastníkmi štúdie a ich meranými médiami.

Multiplatform

Výsledky **gemiusAudience™** sú poskytované zvlášť pre platformu PC (počítače), platformu Phone (mobilné telefóny) a platformu Tablet (tablety). Všetky metriky sú vypočítané na dátach zodpovedajúcich podmnožine populácie internetových užívateľov pre danú platformu. Sociodemografické dáta o profile užívateľov stránok sú dostupné iba pre PC platformu. Dáta sú užívateľom dostupné mesačne v analytickej aplikácii gemiusExplorer. V online rozhraní na adrese aimmonitor.sk, sú základné metriky (RU, PV, V...) aktualizované na dennej báze, vždy za predchádzajúci deň.

Identifikácia užívateľov a výpočet RU

Pre účely výskumu **gemiusAudience™** je internetový užívateľ definovaný ako osoba, ktorá aspoň raz mesačne konzumuje obsah alebo služby dostupné na internete.

Z technického hľadiska meraná aktivita užívateľov na auditovaných webových stránkach vzniká a je agregovaná daným **browserID™**, prítomným v čase, keď sa aktivita odohráva. **BrowserID™** je unikátna identifikačná metóda vyvinutá spoločnosťou Gemius, ktorá sa zakladá na ukladaní užívateľovho ID pre daný

prehliadač v jeho lokálnom úložnom priestore, rovnako ako aj štandardnom cookie priestore. Vo svojej podstate **browserID™** slúži k presne rovnakému účelu ako 3rd party cookies a umožňuje identifikovať aktivitu toho istého užívateľa (s použitím jediného prehliadača) na rôznych internetových stránkach všetkých účastníkov štúdie. Jeho výhoda spočíva vo vyššej miere perzistencie (nižšia miera mazania a opt-out o 2 - 20 %) v porovnaní so štandardnými cookies a v prekonaní problému predvoleného nastavenia o neprijímaní 3rd party cookies v mnohých moderných prehliadačoch, a to najmä v prípade mobilných zariadení.

Výskum **gemiusAudience™** nekončí len pri poskytnutí počtu unikátnych prehliadačov, ale ide ešte ďalej a odhaduje presný počet skutočných užívateľov internetu, ktorí sa za nimi skrývajú. Použitá metóda je založená na predpoklade, že je možné definovať skupinu identifikátorov prehliadača, ktoré by boli reprezentatívne pre všetky identifikátory (či už cookie súbory alebo **browserID™**), ktoré by mohli byť použité pre výpočet reachu média.

Tým, že poznáme priemerný počet zobrazení stránky pre identifikátor v tejto reprezentatívnej skupine a poznáme celkový počet zobrazení stránky vykonaných na médiu, vieme odhadnúť počet identifikátorov, ktoré by boli registrované pre konkrétne médium, ak by nedošlo k žiadnemu vymazaniu identifikátora. V dôsledku toho možno vypočítať reach takéhoto média a teda počet reálnych užívateľov.

Traffic generovaný touto osobitnou skupinou identifikátorov (dobré **browserID™**) by mal mať rovnaké charakteristiky ako traffic generovaný súčtom všetkých identifikátorov. Reprezentatívnosť použitého algoritmu **BEAST™** vo výskume **gemiusAudience™** je dosiahnutá zavedením pravidla, že identifikátory na to, aby patrili do vybranej skupiny dobrých identifikátorov, musia existovať počas celého skúmaného mesiaca, do obdobia ktorého patrí. To znamená, že existovali tak pred, ako aj po skončení tohto obdobia v rozsahu mesiaca a napokon odráža definíciu internetového užívateľa.

Aby sme nemuseli čakať celú dĺžku mesiacov, na to, aby sme na základe zozbieraných informácií preverili presnú kompozíciu skupiny dobrých identifikátorov, vo výskume **gemiusAudience™**, **BEAST™** algoritmus obsahuje vhodný analytický model, ktorý s vysokou istotou určuje, či daný identifikátor patrí alebo nepatrí do skupiny dobrých identifikátorov, na základe charakteristík správania daného **browserID™**.

V tomto modeli je každé **browserID™** popísané dvoma parametrami:

- množstvo času, ktoré uplynulo od poslednej zaznamenatej činnosti v analyzovanom mesiaci do konca mesiaca;
- miera aktivity **browserID™** v priebehu štyroch týždňov predchádzajúcich poslednej zaznamenatej aktivite.

Podľa týchto parametrov sú všetky identifikátory rozvrstvené do podskupín (vrstvy). Pre každú vrstvu sa na základe historických dát vypočíta váha. Tieto váhy predstavujú proporciu "dobrých" identifikátorov v každej

vrstve. Pravdepodobnosť, že identifikátor prehliadača je "dobrý" **browserID™** sa vypočíta na základe pravdepodobnostnej funkcie odvodenéj z historických vstupných dát v dĺžke 6 mesiacov.

Algoritmus použitý vo výskume **gemiusAudience™** na vypočítanie počtu Reálnych užívateľov pre dané médium, možno zhrnúť takto:

- Výpočet "dobrých" **browserID™** a zobrazení stránok nimi generovaných na auditovanom médiu.

$$DobréBrowserID(W) = \sum_{T_1=0}^{30} \sum_{T_2=0}^{28} Pravdepodobost'(T_1, T_2) \cdot BrowserID(W, T_1, T_2)$$

$$DobréPageViews(W) = \sum_{T_1=0}^{30} \sum_{T_2=0}^{28} Pravdepodobost'(T_1, T_2) \cdot PageViews(W, T_1, T_2)$$

kde,

T_1 – je množstvo času, ktoré uplynulo od poslednej zaznamenanej činnosti v analyzovanom mesiaci do konca mesiaca (uvedené v počte kalendárnych dní)

T_2 – je miera aktivity cookie v priebehu štyroch týždňov (28 dní) predchádzajúcich poslednej zaznamenanej aktivite.

- Výpočet počtu identifikátorov prehliadača, ktoré by boli zaregistrované pre auditované médium, keby nenastalo žiadne mazanie identifikátorov (skutočný počet prehliadačov)

$$Prehliadače(W) = PageViews(W) \cdot \frac{DobréBrowserID(W)}{DobréPageViews(W)}$$

čo môže byť zovšeobecnené na celý internet

$$Prehliadače(Internet) = PageViews(Internet) \cdot \frac{DobréBrowserID(Internet)}{DobréPageViews(Internet)}$$

- Výpočet relatívneho reachu auditovaného média v danom časovom období.

$$Reach(W) = \frac{Prehliadače(W)}{Prehliadače(Internet)}$$

To umožňuje vypočítať počet Reálnych užívateľov, ktorý má byť započítaný a obsiahnutý vo výsledkoch výskumu **gemiusAudience™**.

$$ReálniUžívateľia(W) = Reach(W) \cdot InternetováPopulácia$$

Informácie o populácii internetových užívateľov (internetová populácia) sa zhromažďujú použitím prieskumu externej inštitúcie. Použitie algoritmu **BEAST™** vo výskume **gemiusAudience™** umožňuje výpočet oficiálnych výsledkov na nasledujúci deň po skončení obdobia, na ktoré sa výskum vzťahuje.

Nábor panelistov

Užívatelia internetu sú regrutovaní do panelu **gemiusAudience™** výskumu prostredníctvom metodológie CAWI. Dotazníky sú emitované náhodne jednotlivcom z množiny užívateľov, ktorí navštevujú auditované médiá, na základe stanovených podmienok správania a frekvencie. Konfigurácia zahŕňa:

- Identifikovaný obsah auditovaného média,
- Prvotný dátum vytvorenia identifikátora prehliadača,
- predchádzajúce vystavenie dotazníkom,
- dĺžka aktuálnej session,
- homogenita cesty užívateľa,
- pop-up alebo pop-under nastavenie,
- geolokácia,
- výpadky dotazníka,
- frekvencia emisie dotazníka v rámci skupiny oprávnených identifikátorov prehliadača

Po vyplnení dotazníka sa daný užívateľ stáva členom cookie panelu, až do vymazania **browserID™**, s ktorým bol prehliadač zapísaný.

Dotazník obsahuje opt-out možnosť pre užívateľov, ktorí sa rozhodli, že sa do panelu nezapoja a neželajú si byť vystavení zobrazovaniu dotazníka v budúcnosti.

Zber dát

Hlavným zdrojom pre dáta použité v **gemiusAudience™** je site-centric meranie. Site-centric meranie sa vykonáva za použitia metódy „in code tagging“ na základe javascriptu a infraštruktúry pre zber dát spoločnosti Gemius. Zhromaždené site-centric dáta sú okamžite k dispozícii v rozhraní **Prism™** a v obmedzenej forme na www.aimmonitor.sk

Vydavateľ, ktorému patrí obsah stránok nasadzuje meranie v podobe skriptov v nasledujúcom formáte:

```
<script type="text/javascript">
<!--/--><![CDATA[//><!--
var pp_gemius_identifier = 'IDENTIFIKATOR';
// lines below shouldn't be edited
function gemius_pending(i) { window[i] = window[i] || function() {var x = window[i+'_pdata'] =
```

```
window[i+'_pdata'] || []; x[x.length]=arguments;});  
gemius_pending('gemius_hit'); gemius_pending('gemius_event'); gemius_pending('pp_gemius_hit');  
gemius_pending('pp_gemius_event');  
(function(d,t) {try {var  
gt=d.createElement(t),s=d.getElementsByTagName(t)[0],l='http'+((location.protocol=='https:')?'s:');  
gt.setAttribute('async','async');  
gt.setAttribute('defer','defer'); gt.src=l+'//gask.hit.gemius.pl/xgemius.js'; s.parentNode.insertBefore(gt,s);}  
catch (e) {}})(document,'script');  
//--><![>  
</script>
```

Kontakt daného užívateľa internetu s auditovaným obsahom alebo službou spustí hlavný merací skript prítomný na serveri zodpovednom za zber dát (**hitcollector**). Výsledkom interakcie je vytvorenie alebo aktualizácia súčasného identifikátora prehliadača užívateľa a generované informácie o aktivite užívateľov (**hit**) sú odoslané na spracovanie.

Odoslané informácie sa skladajú z dostupných údajov nachádzajúcich sa v prehliadači o zariadení a systéme, správaní sa užívateľa a klasifikácie podporujúcej ďalšiu agregáciu a spracovanie dát na účely **gemiusAudience™** štúdie.

Patrí medzi ne verzia skriptu, identifikátor skriptu, parameter nesúci informáciu o identifikácii iframe, verzia Flash Player, informácia o časovom pásme užívateľa, URL adresa a referer, rozlíšenie obrazovky, veľkosť okna, identifikátor prehliadača, 3rd party cookie identifikátory, technické parametre určujúce emisiu dotazníka, popis akcie, ktorú užívateľ vykonal (napr. zobrazenie stránky) a spôsob, akým boli odoslané dáta vytvorené, IP adresa a user agent.

Overovanie dát a ich filtrovanie

Zozbierané dáta sa ukladajú v databázach spoločnosti Gemius, no nie sú použité okamžite v **gemiusAudience™** výskume. Informácie sú ďalej spracovávané, overované a filtrované.

Gemius vedie aktualizovaný zoznam robotov a crawlers, ktorých činnosť generuje traffic na obsahu a službách účastníkov štúdie, no tento traffic nemôže byť braný ako ľudské správanie. Dáta prichádzajúce z takéhoto zdroja v site-centric štúdii sa zhromažďujú, sú však uchovávané oddelene a nie sú použité pri ďalšom spracovaní.

Filtrovanie kritérií v tejto fáze je nastavené na základe preddefinovaných identifikátorov alebo IP rozsahu a odohráva sa na úrovni **hitcollectoru**. Všetci účastníci štúdie môžu požiadať o rozšírenie definície filtrovania podľa robotov a crawlers ich vlastnej výroby, alebo tých, o ktorých vedia. Počas tohto procesu nedôjde k strate dát. Avšak z hľadiska výskumu **gemiusAudience™** a rozhrania **Prism™**, sa táto časť zaslaných informácií nezobrazuje ani sa nezapočíta.

Po tom, ako sa vyfiltruje non-human generovaný traffic z **gemiusAudience™** výskumu sa spúšťa jej výpočet. Proces filtrovania extrémnych identifikátorov prehliadača je zameraný na identifikáciu a elimináciu pomocou vzorcov správania sa časti zhromaždených dát, ktoré by nemohli byť vytvorené jedinou osobou, avšak boli zhromaždené, identifikované a naakumulované jediným prehliadačom. Behaviorálne obmedzenia uplatňované pri filtrovaní, ktoré vytvárajú non-human aktivitu zahŕňajú:

- maximálne množstvo aktivity, ktorá môže byť generovaná ľudským užívateľom;
- maximálna frekvencia generovania činnosti v meranom časovom období;
- maximálne strávený čas.

Filtrovanie je vykonávané na denných setoch dát a non-human informácie sú odstránené z **gemiusAudience™** vstupu iba z daných analyzovaných setov. To umožňuje presné odstránenie chybných behaviorálnych dát bez potenciálnej straty reálnych dát. To sa môže týkať prípadov s dočasnými prekážkami alebo fraudov, ktoré netrvali počas celého mesiaca.

V ďalšom kroku sú dáta výskumu **gemiusAudience™** očistené od všetkých dát prichádzajúcich z meracích skriptov vložených do iframe, ktoré násobia informácie o jednotlivej akcii internetového užívateľa. Tieto údaje sú generované buď v dôsledku ľudskej chyby pri nasadzovaní merania alebo vedomých akcií zameraných na obchádzanie správnosti merania. Filtrovanie iframe odstráni nadmerné dáta alebo krížové zdieľanie dát medzi viacerými médiami – to je situácia, v ktorej je informácia o akcii užívateľa zaslaná do viacerých účtov účastníkov štúdie v rovnakom čase.

Nakoniec sú vstupné dáta pre výskum **gemiusAudience™** filtrované podľa povolených domén. Cieľom tohto kroku spracovania údajov je zabezpečiť, aby zapísané informácie o aktivite užívateľov internetu pochádzali z obsahu a služieb, ktoré patria ku konkrétnemu vydavateľovi, definovanému súborom subdomén a domén (zvyčajne jednou doménou), ktoré patria danému účastníkovi štúdie.

Ďalším aspektom, ktorý je overovaný pred uskutočnením výpočtu výskumu **gemiusAudience™** sú údaje panelistov získané z dotazníkov. Každá odpoveď účastníkov panelu je overovaná súborom logických pravidiel. Panelisti, ktorých činnosť bola pozastavená a nespĺňajú definíciu internetového užívateľa, rovnako ako aj tí, ktorých odpovede nie sú logicky koherentné sú odstránení zo skupiny.

Výsledky výskumu

Výsledky výskumu **gemiusAudience™** sú pripravované za obdobie jedného mesiaca a zahŕňajú aj denné a týždenné hodnoty metrík. Celá overená dátová sada je potom rozdelená podľa multi-platform prístupu. Toto riešenie zohľadňuje spôsoby, akým používatelia konzumujú auditovaný obsah a služby – typ použitého zariadenia, ktoré sú zoskupované do jednotlivých platforiem.

Takýto prístup zavádza priamy vzťah medzi vzorcami správania sa jednotlivých užívateľov konzumujúcich obsah a služby, bez ohľadu na pôvodný vydavateľom určený dizajn stránok - rozdelenie do mobilných a ne-mobilných webových stránok a služieb. Multi-platform prístup použitý vo výskume **gemiusAudience™** zohľadňuje fakt, že aj s dynamickým responzívnym dizajnom stránok, určitý počet mobilných stránok môže byť zobrazený internetovým užívateľom na svojich počítačoch a naopak, klasické webové stránky je možné zobrazovať na mobilných zariadeniach.

Vo výskume **gemiusAudience™** sú namerané dáta generované užívateľmi internetu na mobilných zariadeniach oddelené od dát z PC platformy. Všetky metriky sú vypočítané na dátach zodpovedajúcich podmnožine populácie internetových užívateľov.

Dodávanie výsledkov výskumu - dáta PC platformy:

- denné, týždenné a mesačné na www.aimmonitor.sk dodané na druhý deň do 12.00 hod po skončení daného časového obdobia
- denné, týždenné a mesačné dáta slovenských užívateľov s ich sociodemografiou dodané v prvý pracovný deň po 15. dni mesiaca nasledujúcom po meranom mesiaci v súbore .gem pre analýzu v aplikácii gemiusExplorer

Dodávanie výsledkov výskumu - dáta platformy Phone:

- denné, týždenné a mesačné na www.aimmonitor.sk dodané na druhý deň do 12.00 hod po skončení daného časového obdobia

Dodávanie výsledkov výskumu - dáta platformy Tablet:

- denné, týždenné a mesačné na www.aimmonitor.sk dodané na druhý deň do 12.00 hod po skončení daného časového obdobia

Výsledky výskumu sú v aplikácii gemiusTraffic/gemiusPrism dostupné online s trojhodinovým oneskorením od poslednej akcie, ktorú zaznamená merací skript (napr. zobrazenie a iné aktivity)

Dáta v online rozhraní na www.aimonitor.sk sú voľne prístupné verejnosti. Dáta dodané vo formáte .gem budú dostupné pre médiá zapojené vo výskume alebo pre agentúry a iné subjekty odoberajúce dáta prostredníctvom knižnice v aplikácii gemiusExplorer alebo prostredníctvom linky na stiahnutie. Typ licencie účastníkov štúdie k jednotlivým dátam určí zadávateľ merania.

Doplňujúce informácie

Navrhovaná Metodika výskumu AIMmonitor sa môže meniť po dohode IAB Slovakia a Realizátora/Gemius Slovakia, pričom podlieha schváleniu Metodickej komisie IAB Slovakia. V prípade zmeny Metodiky nové znenie sa stane záväzným jej zverejnením na webovom sídle IAB Slovakia.

Rozdelenie platformy Mobile na platformy Phone a Tablet

Dňa 3.10.2017 bolo Metodickou komisiou združenia pre internetovú reklamu IAB Slovakia odsúhlasené rozdelenie platformy Mobile na platformy Phone a Tablet. Zmena bude aplikovaná do online aplikácie OLA od 1.12.2017 a v .gem súboroch za mesiac december 2017. Dáta sa nebudú prepočítavať späť. Zmena sa týka taktiež platformy PC, z ktorej sa vylúčia dáta kategorizované ako "Other" a "TV", aby aktuálne názvy platforiem reflektovali ich skutočný obsah. Platforma "PC" bude obsahovať iba traffic generovaný z PC, platforma "Phone" traffic generovaný mobilnými telefónmi a platforma "Tablet" traffic generovaný tabletmi.

Tento metodický dokument bol zaktualizovaný dňa 20.11.2017.