



Metodika výskumu IABmonitor

monitor.iabslovakia.sk

Apríl 2019

1 Úvod

Tento dokument je základnou verziou metodiky projektu IABmonitor – auditovaného výskumu internetovej návštevnosti na Slovensku.

Materiál popisuje základné princípy a parametre projektu, ktoré vychádzajú z dohôd medzi realizátorom projektu a zadávateľom projektu a samotnou metodikou projektu IABmonitor.

1.4.2019 došlo k aktualizácii metodiky na metodiku gemiusOverNight ako aj zmene názvu merania z AIMmonitor na IABmonitor.

1.1 Realizátor projektu

Realizátorom projektu IABmonitor je medzinárodná spoločnosť Gemius S. A. a jej dcérska spoločnosť Gemius Slovakia s.r.o. (www.gemius.com).

Gemius je medzinárodná výskumná a konzultačná agentúra, európsky líder v oblasti prieskumu internetu. Spoločnosť pôsobí v 30 štátoch Európy a Blízkeho východu. Gemius je inovátorom v oblasti hybridnej metodológie merania návštevnosti internetu, ktorá využíva kombináciu dát pochádzajúcich z výskumného panelu a výskumu site-centric, čo poskytuje veľmi spoľahlivé výsledky médiám a media plannerom. Výskum nesie meno gemiusAudience, na Slovensku je prezentovaný pod menom IABmonitor.

1.2 Zadávateľ a vlastník dát

Zadávateľom výskumu IABmonitor a vlastníkom dát je Interactive Advertising Bureau Slovakia (IAB Slovakia). IAB Slovakia je záujmovým združením pôsobiacim v oblasti internetu. V súčasnosti má 39 členov, ktorými sú najväčší slovenskí vydavatelia, mediálne agentúry a ďalší významní hráči slovenského digitálneho trhu. Prevádzkuje jednotný nezávislý auditovaný rebríček návštevnosti slovenských internetových médií IABmonitor. Vo svojej činnosti sa združenie zameriava na štandardizáciu digitálneho reklamného trhu zavedením jednotných pravidiel a štandardov pre internetovú reklamu, ako aj etických pravidiel v digitálnom odvetví. Podporuje ďalší rozvoj a napredovanie slovenského online ekosystému svojimi vzdelávacími aktivitami. Zabezpečuje informovanosť o dianí v segmente digitálneho trhu na Slovensku a vo svete, monitoruje a informuje o zmenách v legislatíve, ktorá je úzko prepojená s digitálnym prostredím. Zároveň funguje ako platforma pre vzájomnú komunikáciu medzi subjektmi podnikajúcimi v oblasti internetovej reklamy a komunikačný kanál s ostatnými profesnými organizáciami v oblasti internetu či orgánmi štátnej správy. V digital odvetví pôsobí od roku 2005 (v tom čase Asociácia internetových médií). Je členom asociácie IAB Europe.

2 Site-centric audit

Výskum návštevnosti IABmonitor je vykonávaný pomocou hybridnej metódy merania, kde sa vychádza tak z meraní na strane webu (site-centric) pomocou javascriptových tagov, tak z meraní na strane užívateľa (user-centric) za pomoci výskumného panelu. Panelom sa rozumejú užívatelia, ktorí na niektorom zariadení, z ktorého prístupujú k internetu, vyplnili regrutačný dotazník (stávajú sa členmi tzv. **Pop-up panelu**). V rámci výskumu je následne monitorovaná aktivita týchto užívateľov naprieč webmi zapojenými do výskumu.

Internet je médium, v rámci ktorého je možné sledovať akcie užívateľa v rámci návštevy stránky alebo používania aplikácie s takmer 100% spoľahlivosťou. Takúto mieru presnosti je možné dosiahnuť vďaka tagom (sledovacím skriptom) vloženým priamo do HTML kódu webových stránok. Tak je možné zaznamenať každú návštevu stránky, rovnako ako všetky súvisiace parametre tejto akcie. Tento spôsob sledovania je pre internetového užívateľa transparentný, nijako užívateľa neobmedzuje a je vykonávaný automaticky. Dáta sú zozbierané od všetkých návštevníkov bez ohľadu na miesto prístupu k internetu (doma, v práci, v kaviarni, v škole a ďalšie) a bez ohľadu na použité zariadenie (stolný počítač, notebook, mobilný telefón, tablet alebo ďalšie zariadenie s povoleným JavaScriptom).

To znamená, že pre získanie počtu zobrazených a navštívených webových stránok a na nich stráveného času, nie je potreba robiť akékoľvek odhady na základe vzorky alebo používať deklaratívne dáta. Namiesto toho môžu byť informácie o návštevníkoch IAB Slovakia registrovaných webových stránok získané so 100% presnosťou vďaka použitiu site-centric merania.

Metóda výpočtu reachu webovej stránky so započítaním nepresností spojenými s mazaním BrowserID je popísaná v nasledujúcich kapitolách.

Presné hodnoty počtu page views, návštev webových stránok, na nich strávenom čase a počtu návštevníkov (Reálni užívatelia) je možné získať len pre webové stránky merané pomocou site-centric auditu.

Kombinácia dát externej štúdie a site-centric auditu je použitá k nastaveniu panela určenej váhy pre každého reportujúceho panelistu.

2.1 Identifikácia užívateľa

Gemius vyvinul a v SR implementoval (podobne ako vo väčšine ostatných krajín, kde je aktívny) unikátnu metódu identifikácie používateľa označovanú ako BrowserID (BID), ktorá je založená na ukladaní ID používateľa pre daný prehliadač v jeho Local Storage a zároveň v klasickej Cookie tretej strany, čo je spôsob, ktorým boli užívatelia identifikovaní pred zavedením BID. Jeden BID identifikuje práve jeden prehliadač. Výhodou BrowserID je jeho väčšia odolnosť v porovnaní s Cookies voči jeho mazaniu a v riešení problému s neprijímaním identifikátorov Cookies tretej strany hlavne u mobilných zariadení. Jeho vyššia rezistencia tak pozitívne prispieva k lepším výsledkom algoritmu výpočtu Reálnych Užívateľov pre všetky platformy a zároveň k presnejšiemu počítaniu ukazovateľov návštev a stráveného času.

2.2 Meracie skripty

Dáta týkajúce sa návštevnosti zapojených webových stránok a streamovaného obsahu a aplikácií sú získavané pomocou skriptov (meracích kódov) gemiusPrism vložených do meraných webových stránok, videa alebo mobilných aplikácií.

Gemius meracie skripty sú “website-friendly”:

- pre prevádzkovateľa sa jedná o transparentné meranie, žiadny zásah z jeho strany nie je pre zaistenie merania potrebný
- rešpektujeme súkromie internetových užívateľov a ich právo odmietnuť účasť v štúdiu
- meranie je realizované nezávisle na internetovom prehliadači, OS a zariadení, ktoré užívateľ používa
- meranie žiadnym spôsobom neovplyvňuje užívateľskú skúsenosť na webových stránkach alebo aplikáciách, skripty pracujú v asynchrónnom móde a nerušia tak načítavanie obsahu
- hypotetický výpadok zberného skriptu neovplyvní načítanie webových stránok, videa alebo aplikácie, keďže skripty sú optimalizované pre webové stránky, mobilné stránky, mobilné aplikácie, desktopové aplikácie a ďalšie
- skripty podporujú zabezpečené pripojenie (HTTPS / SSL) a RIA (napríklad Flash alebo Ajax) pre meranie akcií.
- skript zbiera informácie o FRAME alebo IFRAME, v ktorom je umiestnený. To umožňuje zisťovať, či sú stránky správne skriptované a či nedochádza k navyšovaniu návštevnosti. Problematické hity je následne možné odfiltrovať.

Pravidlá pre skriptovanie stránok sú rovnaké pre všetkých členov IAB Slovakia. Základné pravidlá sú nasledujúce:

- Stránky zúčastňujúce sa merania vkladajú skripty do stránok na základe inštrukcií od realizátora.
- Každá webová stránka môže obsahovať len jeden merací tag, pokiaľ nie je dohodnuté ináč realizátorom. Pokiaľ bude do stránky vložených viac meracích skriptov, bude započítaný len jeden hit. Ostatné hity budú automaticky zaslané do dátového úložiska nesprávnych hitov. Prevádzkovateľ je povinný nadbytočné tagy zo stránok odstrániť.
- Prevádzkovateľ nesmie žiadnym spôsobom modifikovať meracie skripty. Len skripty poskytnuté realizátorom môžu byť použité pre daný web.
- Na daný web sa môžu nasaďiť len skripty, ktoré sú pre daný web určené.
- Site-centric skriptami nie je povolené monitorovať reklamný obsah a pop-up okná, ktoré sú zobrazené bez akcie užívateľa.

Podrobnejšie pravidlá môžu byť pripravené v spolupráci s IAB Slovakia, tak aby meranie zodpovedalo požiadavkám transparentného merania a aby bolo ošetrené meranie zvláštnych udalostí (napr. Flash, hry, desktop aplikácie, iFramy, autorefresh).

Neplnenie pravidiel zo strany médií môže viesť k odstráneniu média z merania.

Site-centric meranie, založené na zbere štatistík v rámci navštívených stránok, je realizované vďaka spojeniu medzi webovým prehliadačom užívateľa a serverom, ktorý dáta zbiera. To všetko pomocou meracieho skriptu umiestneného do zdrojového kódu.

Skript gemiusPrism sa skladá z 2 častí:

- **hlavný skript gemius.js**
- **merací skript**, ktorý musí byť vložený do HTML kódu meranej stránky

Merací skript vyzerá nasledovne:

```
<script type="text/javascript">
<!--><![CDATA[<!--
var gemius_identifier = 'IDENTIFIER';
// lines below shouldn't be edited
function gemius_pending(i) { window[i] = window[i] || function() {var x = window[i+'_pdáta'] =
window[i+'_pdáta'] || []; x[x.length]=arguments;}};
gemius_pending('gemius_hit'); gemius_pending('gemius_evont'); gemius_pending('pp_gemius_hit');
gemius_pending('pp_gemius_evont');
(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+'://PREFIX.hit.gemius.pl/gemius.js'; s.parentNode.insertBefore(gt,s);} catch
(e) {})(document,'script');
--><![>
</script>
```

Kde: **IDENTIFIER** je unikátny identifikátor projektu

PREFIX je prefix servera (nazývaného hitcollector), ktorý meria traffic na monitorovanej webovej stránke

Pokiaľ je monitorovaná stránka zobrazená, vykoná sa i merací skript a pošle premennú gemius_identifier na hlavný gemius.js skript. Keď sa vykoná hlavný skript, je na server Gemiusu poslaná požiadavka na zobrazenie transparentného, 1x1 pixelu veľkého, obrázku. Takto systém zaznamená vykonanú udalosť (event).

Ukážka požiadavky na Gemius server (hitcollector) vyzerá nasledovne:

```
http://pre.hit.gemius.pl/_1378989199803/redot.js?!=81&ltime=105&lzdáta=KIYVMEheo5yGBpn3IZ_DxqQGq
BavPuXxunFiXB0p25j.Y7/1372843602640/&fpdáta=hs93FOt6sAMxnqYOcKLu67RuhVghRVSlIPdP2y4qfU..
D7&vis=1&id=ofJF1ju49R59SkiRvVb49LS4.K5yalx3mWeJu9eRR_7.G7&fr=1&tz=-120&fv=Shockwavo Flash
11.7 r700&href=http://biuro.gemius.pl/~akozak/&ref=&screen=1280x1024r1000&col=24&window=1280x811
```

kde: <http://pre.hit.gemius.pl> – príklad adresy servera, na ktorý sa webový prehliadač pripája
1378989199803 – kódovaný dátum a čas (timestamp)
redot.js – 1x1 px obrázok

Hitcollector zbiera dáta o trafficu na meranej webovej stránke a posiela ich do Gemius data processing centra. Dáta zbierané systémom gemiusPrism neobsahujú žiadne dôverné informácie o návštevníkovi.

Gemius meracie skripty zbierajú nasledujúce informácie:

- Skript identifikátor
- Informácie o tom či je skript vložený do iFrame
- Flash Player verzia
- Časové pásmo užívateľovho počítača vo vzťahu k GMT (v minútach)
- URL adresa, na ktorej bol skript vykonaný
- URL adresa odkazujúcej webovej stránky
- Rozlíšenie obrazovky (v pixeloch)
- Veľkosť okna internetového prehliadača (v pixeloch)
- Tabuľka s voliteľnými parametrami klienta (voliteľné)
- Local storage identifikátor
- Cookie prvej strany / BID identifikátor
- Visibilita okna prehliadača
- Typ udalosti (Event)

Nasledujúce parametre sú získané z HTTP hlavičky:

[IP]	IP adresa, z ktorej bolo vykonané pripojenie (posledné 4 bity sú vynechané)
User – Agent	Umožňuje rozpoznať typ použitého zariadenia, jeho operačný systém a webový prehliadač
X-OperaMini-Phone-UA	Parameter, ktorý pomáha rozpoznať typ mobilného zariadenia
Accept-Language	Preferovaný jazyk v prehliadači používateľa
Referer	URL adresa, kde bol skript vykonaný
X-Forwarded-For:	Pôvodná IP adresa prevedená proxy serverom (anonymizovaná zmazaním posledných 4 bitov v binárnom zobrazení IP adresy)
Client-ip	Pôvodná IP adresa prevedená proxy serverom (anonymizovaná zmazaním posledných 4 bitov v binárnom zobrazení IP adresy)
Cookie	Identifikátor používateľa

2.3 Čistenie dát

Základom dát o návštevnosti sú údaje zo site-centric merania. Aby bolo možné zaistiť spoľahlivé a porovnateľné meranie naprieč stránkami, sú na dáta prostredníctvom gemiusAudience aplikované nasledujúce štandardy a kontrolné postupy:

- Pre výpočet socio-demografie je akceptovaná len domáca návštevnosť (vygenerovaná užívateľmi v Slovenskej republike). Návštevnosť zo zahraničia je odfiltrovaná.

- Na meranie nemá vplyv, či je kód typu site-centric umiestnený do kódu webovej stránky dvakrát či viackrát (namiesto len jediného umiestnenia). Kódy spoločnosti Gemius automaticky odhalia prípady, kedy je merací kód umiestnený viac než raz a dáta sú započítané len raz.
- Odfiltrovaná je tiež návštevnosť v prípade automaticky sa obnovujúcich (refreshujúcich) stránok.
- Odfiltrovaná je z dát tiež návštevnosť v prípade nesprávne vložených sledovacích kódov do IFRAMES a bannerov.
- Akceptovaná je len návštevnosť z povolených webových domén.
- Sú aplikované pravidlá pre filtrovanie extrémnej aktivity.
- Možnosť filtrovania interného traffícu generovanom spoločnosťou, ktorej médium je súčasťou štúdie gemiusAudience

Ďalšie filtrovanie spočíva v odstránení aktivity **extrémnych BrowserID**, ktoré generujú aktivitu tak veľkú, že nemôže byť vytvorená jedným človekom. Aktivita každého BrowserID je hodnotená na dennej báze a pokiaľ je vyhodnotená ako extrémna, je BrowserID z výpočtov vyradené. Aktivita za tieto dni nie je zohľadnená ani v prípade denných a mesačných dát (aktivita extrémnych BrowserID je vyradená z tých dní, kde bola aktivita vyhodnotená ako extrémna).

Denná aktivita BrowserID je vyfiltrovaná pre nasledujúce BrowserID:

- BrowserID, ktoré bolo aktívne len na jednom serveri a vygenerovala tam viac ako 1000 PV za jeden deň.
- BrowserID, ktoré na internete strávilo viac než 20 hodín daný deň.
- BrowserID, ktoré v priemere vygenerovalo viac ako 2 PV za sekundu. Pokiaľ BrowserID vygenerovalo v danom dni viac než 5000 PV, je frekvencia znížená na PV za sekundu.

2.4 Rozdelenie návštevníkov podľa použitého prístupového zariadenia

Gemius aktuálne reportuje dáta rozdelené do dvoch hlavných platforiem - PC a Mobil, ktoré sa ďalej delia na ďalšie platformy:

1. PC
 - a) PC Home
 - b) PC Work
2. Mobil
 - a) Mobilné telefóny
 - b) Tablety

Informácie o delení platformy PC na PC Home a PC Work je v kapitole 8, PC Home a PC Work.

Identifikácia zariadenia je vykonávaná pomocou HTTP hlavičky, ktorá je jedným z parametrov zbieraným meracími skriptami. Na základe tejto identifikácie je každé zobrazenie stránky priradené príslušnej platforme. Gemius udržiava databázu niekoľko tisíc zariadení, vrátane tabletov, smartphonov a ďalších.

Podiel iných platforiem (napr. Smart TV) na celkovom trafficu sa momentálne pohybuje v zanedbateľných hladinách z celkového počtu registrovaných zobrazení stránok. Gemius tento podiel kontinuálne sleduje a je pripravený zahájiť s IAB Slovakia diskusiu o pridaní ďalšej platformy do oficiálnych dátových výstupov akonáhle tento podiel prekročí dohodnutú hranicu.

3 Výskum sociodemografie

Výskum sociodemografického profilu návštevníkov aktívne meraných serverov a ich sekcií je v IABmonitor realizovaný na paneli respondentov, tzv. Pop-up panel. Pop-up panel je aktuálne používaný len na platforme PC, pre mobilné platformy bola regrutácia spustená v decembri 2018 a produkcia bude o dáta z tohto panelu rozšírená v priebehu roku 2019.

3.1 Typy panelov

Gemius aktuálne prevádzkuje v rámci IABmonitoru jeden typ panelu – Pop-up panel. Technológia umožňuje aj regrutáciu a použitie tzv. Softvérového panelu pre meranie a reporting návštevnosti médií mimo okruh webov aktívne zapojených do štúdie IABmonitor, avšak táto technológia nie je aktuálne vo výskume používaná.

Pre online panel existujú dve významné podmienky: panel by mal byť reprezentatívny a mal by byť čo najväčší. Je možné, že panel regrutovaný online nemusí byť úplne reprezentatívny. Na druhú stranu, pokiaľ je panel regrutovaný offline, je nepravdepodobné, že dostupná vzorka nazbiera dostatok site-centric dát a umožní dostatočnú mieru analýzy, tak, ako je to požadované odberateľmi dát.

Vzhľadom k tomu, že panel je konštruovaný k meraniu návštevnosti internetu, je možné, aby online regrutácia bola dominantnou metódou regrutácie. Takéto riešenie je cenovo efektívne vzhľadom k možnej veľkosti panelu. Pokiaľ bude regrutácia prebiehať naprieč všetkými internetovými stránkami, potom bude možné zasiahnuť veľkú časť online populácie. Regrutácia prebieha pomocou pop-up dotazníkov.

3.2 Pop-up Panel

Internetový užívateľ, ktorý vyplnil regrutačný dotazník sa stáva Pop-up panelistom. Pop-up panelista má svoj demografický profil naviazaný na BrowserID (BID).

Aktivita Pop-up panelistu sa meria pomocou site-centric skriptov, ktoré sú vložené do html kódu webových stránok.

Site-centric skripty zbierajú tieto dáta: identifikátor stránky (ID), URL stránky, BID, geolokácia na báze IP adresy, prehliadač, operačný systém, referer a ďalšie.

Logy aktivity BID obsahujú:

- Presný čas page view
- číslo Pop-up panelistu
- Identifikátor uzlu v strome médií
- Page view URL

Tento Log aktivity môže byť poslaný programu, ktorý počíta čas návštevy a návštevy internetových užívateľov.

3.2.1 Mobilný Panel

Internetový užívateľ, ktorý vyplnil regrutačný dotazník na mobilnom zariadení sa stáva mobilným panelistom. Mobilný panelista má svoj demografický profil naviazaný na BrowserID (BID). Podmienkou je aby Browser ID (BID) bolo priradené počas celého mesiaca (to znamená priradené pred meraným mesiacom a nevyazané počas celého daného mesiaca).

Aktivita mobilného panelistu sa meria pomocou site-centric skriptov, ktoré sú vložené do html kódu webových stránok.

Site-centric skripty zbierajú tieto dáta: identifikátor stránky (ID), URL stránky, BID, geolokácia na báze IP adresy, prehliadač, operačný systém, referer a ďalšie.

Logy aktivity BID obsahujú:

- Presný čas page view
- číslo Pop-up panelistu
- Identifikátor uzlu v strome médií
- Page view URL

Tento Log aktivity môže byť poslaný programu, ktorý počíta čas návštevy a návštevy internetových užívateľov.

Šablóny mobilných dotazníkov sú navrhnuté tak, aby boli plne kompatibilné s dotykovými obrazovkami. Zoznam odpovedí je skrátený oproti dotazníkom zobrazeným na PC zariadeniach a počet otázok je obmedzený na minimum.

Mobilná verzia dotazníka môže byť zobrazená vertikálne i horizontálne. Aby sa skrátily čas potrebný na dokončenie, dotazník je navrhnutý tak, aby bol správne zobrazený a dobre viditeľný na všetkých druhoch mobilných zariadení.

3.3 Regrutácia do panelov

Účastníci sú do **Pop-up panelu** regrutovaní prostredníctvom dotazníkov, náhodne zobrazovaných na členských webových stránkach IAB Slovakia. Tieto dotazníky poskytujú demografické a iné dáta potrebné k analýzam a váženiu.

Regrutácia do Pop-up panelu prebieha priebežne po celú dobu trvania štúdie. Internetoví užívatelia sú náhodne vyberaní a prostredníctvom dotazníka pozvaní do panelu. Žiadosť o vyplnenie dotazníka je každý deň odoslaná vopred definovanému počtu internetových

prehliadačov, ktoré obsahujú Gemiusom vydané BrowserID (BID). Interval medzi jednotlivými žiadosťami sa mení tak, aby bola neustále udržaná požadovaná veľkosť panelu.

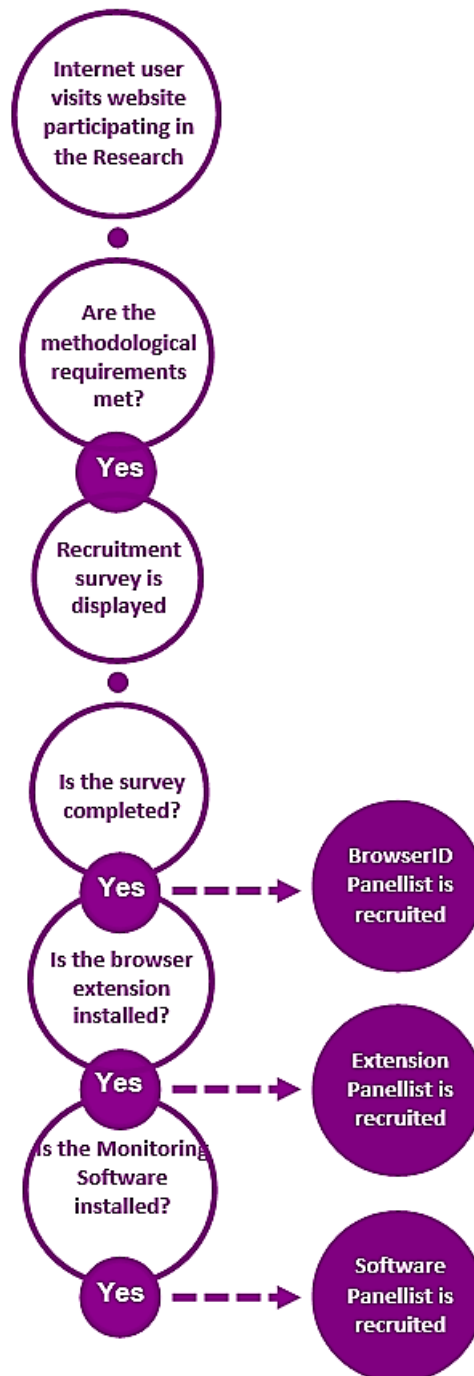
Užívatelia internetu sú regrutovaní do Pop-up panelu prostredníctvom metodológie CAWI. Dotazníky sú emitované náhodne jednotlivcom z množiny užívateľov, ktorí navštevujú auditované médiá, na základe stanovených podmienok správania a frekvencie. Konfigurácia zahŕňa:

- 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 BrowserID 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 vystavovaní zobrazovaniu dotazníka v budúcnosti.

Obrázok 1: Schéma regrutačného procesu



Regrutácia do **Mobilného panela** prebieha podľa obdobných pravidiel. Viac ukazuje schéma nižšie:

Obrázok 2: Schéma regrutačného procesu mobilného panelu



Vzhľadom k vyššie zmieneným pravidlám regrutácie je jedinou podmienkou, aby sa stal respondent členom online panela vrátane nainštalovaného doplnku do prehliadača, akceptácia BID.

Akonáhle jednotlivec vyplnil a odoslal regrutačný dotazník, nie je z jeho strany nutná žiadna ďalšia akcia. Žiadne odmeny nie sú ponúkané, to eliminuje možnú zaujatosť spojenú s ponúkaním peňažných odmien.

Pokiaľ panelista nezmaže svoj BID identifikátor, zostáva v paneli a priebežne sú overované jeho dáta a aktivita.

Do výsledkov štúdie sú zahrnutí len panelisti, ktorí prejdú verifikačným procesom. Cieľom verifikačného procesu je odstrániť z štúdie takých respondentov, ktorí nepatria do populácie sledovanej v rámci štúdie, rovnako ako respondentov, ktorí uviedli nepravdepodobné kombinácie demografických údajov. Procedúra overenia má dve fázy:

- vylúčenie panelistov, ktorí nespádajú do cieľovej populácie štúdie (napr. nie sú obyvatelia SR);

- kontrola parametrov za účelom nájdania a odstránenia tých panelistov, ktorí uviedli nepravdepodobné kombinácie demografických údajov (napr. vek menej ako 15 rokov a samostatné živobytie).

Regrutačný dotazník tiež zahŕňa otázku o počte ľudí, ktorí počítač používajú, a o množstve čase, ktoré panelista strávi na počítači. Odpovede na tieto otázky sú použité pri filtrovaní Pop-up panelistov tak, aby boli zahrnutí len tí panelisti, ktorým bol pridelený BID identifikátor.

Všetky dáta sú ďalej čistené a vážené podľa nasledujúcich zásad:

- Štrukturálna reprezentatívnosť – Externá štúdia je využitá ku kalibrácii panelu podľa pohlavia, veku, úrovne dosiahnutého vzdelania, miesta bydliska, socio-ekonomického štatútu a frekvencie použitia internetu.
- Behaviorálna reprezentatívnosť – objem nameranej návštevnosti všetkých členských webových stránok v rámci IAB Slovakia je použitý k váženiu vzorky aktivity získaného z panelu užívateľov.

Tento proces vyžaduje váženie, ktoré každému panelistovi priradí špecifickú váhu. To ďalej určuje, koľko ľudí v rámci slovenskej internetovej populácie každý panelista reprezentuje.

3.4 Veľkosť panelu

Minimálna veľkosť reportujúceho PC a mobilného panelu IABmonitor je 5 000 panelistov pre každý panel. Pri platnosti limitu 45 panelistov je táto veľkosť dostačujúca pre konštrukciu sociodemografického profilu pre väčšinu meraných uzlov (serverov a sekcií). S rastúcim panelom rastie i počet respondentov na jednotlivých uzloch, čo však neznamená dramatické zlepšenie situácie v počte uzlov pod limitom.

4 Meraná populácia a externá štúdia

4.1 Meraná populácia

Výskum gemiusAudience počíta počet „Reálnych užívateľov“ s využitím štúdie internetovej populácie, site-centric dát a pridaním demografických údajov z panelu.

Vďaka vyžitiu prístupu so „skutočnými užívateľmi“ sú internetové stránky merané metodikou, ktorá pracuje s ľuďmi (a nie s cookies, počítačmi alebo IP adresami), ich zvykmi a sociodemografickými profilmi. Tým pádom je porovnateľná s tradičnými médiami, ktoré tiež svoju obľúbenosť asociujú so skutočnými spotrebiteľmi (a nie s počtom televízií alebo rádii), dosahom, podielom na trhu a cieľovými skupinami.

Meraná populácia zahŕňa všetkých používateľov internetu zo Slovenskej republiky vo veku 12-79 rokov.

Populácia skúmaná v rámci štúdie sa skladá zo slovenských internetových užívateľov, ktorí sa k internetovým stránkam pripájajú z územia Slovenskej republiky a ktorí spadajú do vekového rozmedzia, stanoveného v rámci externej štúdie.

Veľkosť skúmanej populácie na konkrétnej úrovni bude určená ako počet Reálnych užívateľov pre daný uzol a v danom časovom úseku v súvislosti s algoritmom výpočtu Reálnych užívateľov po započítaní mazania BID.

Návštevnosť pre konkrétne sekcie alebo skupiny webových stránok budú počítané na základe stromu médií. Strom médií bude definovaný v súvislosti i s procedúrami popísanými v rámci tohto dokumentu.

4.2 Externá štúdia

Výsledky externej štúdie sú základom štrukturálneho váženého nutného k zaisteniu demografickej reprezentativity.

Cieľom externej štúdie je zistiť rozsah celkovej populácie, voči ktorej budú výsledky štúdie vážené. Štúdia sa predovšetkým snaží zistiť:

- veľkosť populácie internetových užívateľov v Slovenskej republike, naprieč všetkými meranými platformami;
- štruktúru slovenskej populácie používajúcej internet a jej podmnožiny populácie internetových užívateľov v závislosti na mieste prístupu k internetu (doma, v práci, v škole, ďalšie);
- základné prvky správania sa na internete (napríklad frekvenciu celkového používania internetu v závislosti na mieste prístupu k internetu, históriu používania internetu a ďalšie);

Výsledky externej štúdie sú základom štrukturálneho váženého nutného k zaisteniu demografickej reprezentativity.

Hlavným zdrojom dát v súčasnom projekte je Omnibus spoločnosti Focus, ktorý si Gemius objednáva 3x ročne. V rámci štúdie sú dotazované internetové domácnosti (domácnosti, v ktorých má aspoň jeden člen starší 12 rokov aspoň raz mesačne prístup k internetu), ktorých štruktúra je reprezentatívna vzhľadom k oficiálnym dátam internetových domácností v SR.

5 Algoritmus pre reportovanie Reálneho užívateľa (site-centric)

V štúdiu návštevnosti všetkých návštevníkov internetu sa často používa identifikácia návštevníka pomocou jednoznačného identifikátoru – cookie, BID. Keďže často nastávajú situácie, kde je identifikácia pomocou cookies/BID nemožná (užívatelia počítačov nepovolujú ukladanie cookies/BID na svoj počítač, cookies/BID na svojom počítači pravidelne či nepravidelne mažú, reinstalujú systém a tým cookie/BID odstránia, alebo napr. správca lokálnej siete nastaví politiku mazania cookies pre všetkých užívateľov alebo je možné cookies modifikovať na firewalle a tým zabrániť ich ukladaniu). Občas je identifikácia pomocou cookies/BID nedokonalá (jeden počítač používa viac užívateľov, alebo jeden užívateľ používa viac počítačov alebo viac prehliadačov na jednom počítači).

Z toho vyplýva, že počet BID zaznamenaných na meranom serveri pri meraní návštevnosti nie je rovný počtu skutočných užívateľov internetu. Nie je teda možné vierohodným spôsobom stanoviť reach meraných serverov len na základe počtu BID. Preto je v rámci štúdie IABmonitor **implementovaná metodika BEAST**, ktorá na základe informácií o správaní BID stanovuje počet Reálnych užívateľov (RU), teda odhad počtu skutočných ľudí.

Použitá metóda je založená na predpoklade, že je možné definovať skupinu identifikátorov prehliadačov, ktorá by reprezentovala všetky identifikátory a ktoré by mohli byť použité pre výpočet reachu daného média. Pokiaľ poznáme priemerný počet zobrazených stránok jedným BID v tejto reprezentatívnej skupine, a poznáme celkový počet zobrazení stránok na danom médiu, je možné odhadnúť počet identifikátorov, ktoré by boli registrované pre konkrétne médium, pokiaľ by neexistoval problém mazania identifikátorov. Vďaka tomu je možné vypočítať reach média a z neho potom odvodiť počet Reálnych užívateľov.

Aktivita generovaná v rámci tejto zvláštnej skupiny identifikátorov (dobré BrowserID), by mala mať rovnaké vlastnosti ako aktivita generovaná všetkými identifikátormi. Reprezentativita použitého algoritmu BEAST sa dosiahne zavedením pravidla, že identifikátory, ktoré by mali patriť do reprezentatívnej skupiny, musia existovať v celom meranom mesiaci. To znamená, že identifikátory museli existovať pred začiatkom meraného obdobia, ale i po jeho skončení. Aby sa zabránilo čakaniu na koniec celého mesiaca a aby sa preverilo, že identifikátor existoval po celý mesiac, metodika BEAST pracuje na analytickom module, ktorý s veľkou mierou istoty určuje, či daný identifikátor patrí do súboru dobrých BrowserID. V rámci tohoto modulu je každé BrowserID popísané pomocou dvoch parametrov:

- Doba, ktorá uplynula medzi poslednou aktivitou BrowserID a koncom daného mesiaca
- Miera aktivity daného BrowserID v posledných štyroch týždňoch predchádzajúcim poslednej zaznamenanej aktivite

Na základe týchto parametrov sú všetky identifikátory stratifikované do podskupín. Pre každú podskupinu je vypočítaná váha na základe historických dát. Tato váha predstavuje proporčné zastúpenie dobrých identifikátorov v každej podskupine. Pravdepodobnosť, že BrowserID je ‚dobré‘, je vypočítané na základe pravdepodobnostnej funkcie, ktorá je odvodená z historických dát za posledných 6 mesiacov.

Algoritmus, ktorý je používaný v IABmonitore pre výpočet RU, je ten istý pre všetky reportované platformy a má nasledujúcu podobu:

A) Výpočet dobrých BrowserID a PV generovaných týmito ID na zapojených médiách:

$$GoodBrowserID(W) = \sum_{T_1=0}^{30} \sum_{T_2=0}^{28} P(T_1, T_2) \cdot BrowserID(W, T_1, T_2)$$

$$GoodPageViews(W) = \sum_{T_1=0}^{30} \sum_{T_2=0}^{28} P(T_1, T_2) \cdot PageViews(W, T_1, T_2)$$

kde

T1 – počet dní, ktorý uplynul od poslednej zaznamenanej aktivity v analyzovanom mesiaci do konca mesiaca

T2 – počet dní, v ktorých bolo BrowserID aktívne v štyroch týždňoch (28 dní) predchádzajúcich poslednej aktivite.

P(T1, T2) – pravdepodobnosť, že BrowserID z podskupín T1, T2 je ‚dobré‘

BrowserID(W, T1, T2) – počet BrowserID z podskupín T1, T2 aktívnych na stránke W.

PageViews(W, T1, T2) – počet zobrazení stránky generovaných BrowserID z podskupín T1, T2 aktívnych na stránke W.

- B) Výpočet počtu prehliadačov, ktoré by boli zaznamenané na serveri W, pokiaľ by neexistoval problém mazaniu cookies/identifikátorov:

$$Prehliadače(W) = PageViews(Internet) \cdot \frac{GoodBrowserID(W)}{GoodPageViews(Internet)}$$

Čo môže byť zovšeobecnené pre celý internet:

$$Prehliadače(Internet) = PageViews(Internet) \cdot \frac{GoodBrowserID(Internet)}{GoodPageViews(Internet)}$$

- 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.

Algoritmus, ktorý sa v rámci IABmonitor bežne používa na výpočet Reálnych užívateľov, dokáže vypočítať počty používateľov pre konkrétne platformy, ale nemožno použiť na výpočet Reálnych užívateľov za všetky platformy dohromady (ďalej len RU). Na účely výpočtu RU bez ohľadu na geolokáciu a platformu je použitý výpočet založený na predpoklade, že poznáme RU pre jednotlivé platformy (vrátane veľkostí populácií jednotlivých platforiem) a ďalej z pozorovania, akým spôsobom závisí Reach webu na aktivite používateľov.

Podrobný popis tejto metodiky prekračuje rámec tohto dokumentu. Na tomto mieste len uvedieme, že uvedená metóda je postavená na nasledujúcom predpoklade, že pre stránky s malou

frekvenciou návštev je pravdepodobné, že každá nová návšteva zvyšuje Reach navštívenej stránky. S tým, ako rastie frekvencia návštev, klesá rýchlosť rastu Reach pre danú stránku.

6 Príprava sociodemografických dát

Reportujúce panely sa stanovujú pre každú z platforiem PC Home, PC Work, Mobilné telefóny, Tablety. Aby bol panelista vybraný do reportujúceho panelu, musí spĺňať nasledujúce podmienky:

- Je členom panelu aspoň 28 dní.
- Spĺňa podmienky dané externou štúdiou
- Má validný sociodemografický profil. Ak sociodemografický profil neprejde validačným procesom, je panelista odstránený z reportujúceho panelu. Príklad validačného pravidla: Ak je panelista mladší ako 14 rokov a má ukončené vysokoškolské vzdelanie, nie je zaradený do reportujúceho panelu.
- Má kompletný súbor otázok z rekrutačného dotazníka. Otázky, ktoré boli do dotazníka pridané až po regrutácii panelistov má panelista kódované ako "neodpovedal". Panelista môže mať niektoré otázky týmto spôsobom nezodpovedané. V ostatných prípadoch je vyradený z reportujúceho panelu.
- Bol aktívny v danom reportovanom období

Pokiaľ panelista splní tieto podmienky, pracujeme s ním v rámci produkcie dát.

6.1 Štruktúrne a behaviorálne váženie

Pre váženie existujú dva dôvody:

- 1) Štruktúrne váženie má za cieľ kompenzovať rozdiely v štruktúre internetovej populácie a štruktúre panelu z hľadiska sociodemografických premenných.
- 2) Behaviorálne váženie má za cieľ kompenzovať rozdiel medzi nameranou návštevnosťou pomocou site-centric metódy a návštevnosťou panelistov.

V priebehu váženia sú určené hodnoty váh, ktoré umožňujú prispôbiť merané výsledky štruktúre internetovej populácie, konkrétne

- upraví štruktúru panelu tak, aby sa zhodovala so socio-demografickou štruktúrou internetovej populácie Slovenskej republiky. Panel bude prispôbený hodnotám sociodemografických premenných z externej štúdie. Zoznam premenných používaných pre váženie (štruktúrne váženie):
 - pohlavie
 - vek (kategorizovaný)
 - vzdelanie
 - veľkosť miesta bydliska
 - kraj
 - miesto používania internetu (v práci/v škole)

- upraví štruktúru panelu podľa štruktúry návštevnosti internetových stránok, ktoré sa účastnia merania metódou Site-centric. Váhy budú určené tak, aby súčet vážených výsledkov (behaviorálnych premenných) panelistov, ktorí pristúpili na konkrétnu stránku, zodpovedal výsledkom meraniu Site-centric (behaviorálne váženie).

6.2 Kontrola kvality dát

Validácia je proces, pri ktorom je hodnotená kvalita dát nazbieraných v priebehu štúdie a ich hodnota k dosiahnutiu cieľov štúdie. Počas validačného procesu je súbor dát „čistený“, a to odstránením chybných dát, spôsobených pravdepodobne chybou, alebo vedomou manipuláciou štúdie.

Každá internetová stránka je zodpovedná za riadne vloženie site-centric skriptov podľa pokynov realizátora. Zobrazenia generované nedovolenými mechanizmami budú zamietnuté. Reportovaný bude len traffic generovaný slovenskými BID.

Pre overenie logickej konzistencie nazbieraných dát je použitá kontrola vzorky. Pokiaľ sú v priebehu overovania logickej konzistencie niektoré dáta nekonzistentné, sú vyradené z vážiaceho procesu a nebudú zahrnuté do výsledkov merania, dokiaľ nebudú overené a opravené.

V rámci prípravy dát pre produkciu prebieha filtrovanie nežiadúceho trafficu. Filtrovanie extrémnych BID a ich aktivita, ďalej sa filtruje traffic získaný v rámci systémov na automatické generovanie návštevnosti z niektorých iFramov.

7 PC Home a PC Work

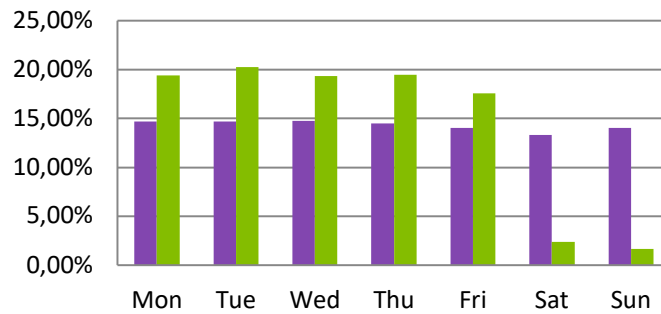
Zavedenie platforiem PC Home a PC Work nám umožňuje rozdeliť prezentované dáta podľa lokácií používania internetu so zachovaním na zreteli, že niektorí užívatelia sa pripájajú z jedného PC napr. len doma alebo len v práci. Iní užívatelia sa ale môžu vďaka laptopu pripájať z jedného PC ako doma tak v práci. Rôzne lokácie pripojenia tiež môžu označovať rôzne vzorce správania. Napr. pre PC home môže byť typická konzumácia video obsahu, kým v rámci návštevnosti z práce môže byť typické využívanie napr. korporátnych webov. Rovnako tak sa pre obe platformy môže líšiť strávený čas s tendenciou tráviť viac času na internete doma.

Platformy Home a Work sú vytvorené na základe toho, že každá IP adresa je priradená do jednej z kategórií Home alebo Work a rovnako tak je do kategórie Home a Work priradený aj každý validovaný panelista. Na základe spojenia týchto platforiem následne vzniká celá platforma PC.

7.1 Klasifikácia IP adres

Prvá fáza klasifikácie IP adres vychádza z údajov z dotazníka, kde panelisti vyplňajú, kde sú pripojení v okamihu vyplňania dotazníka.

Pre tieto IP adresy sa následne vypočítava, kedy sú aktívne (aké dni, aké časti dňa) a ako dlho je každá IP adresa aktívna. Cieľom je identifikovať vzorce aktivity typické pre platformu Home a pre platformu Work. Tieto informácie sú použité v rámci automatického strojového učenia (algoritmus náhodný les), aby bol vytvorený a validovaný model pre priradenie IP adries do kategórie Home alebo Work. Výsledkom je kategorizácia všetkých IP adries z IABmonitor (všetky site-centric dáta) do jednej z kategórií Home alebo Work.



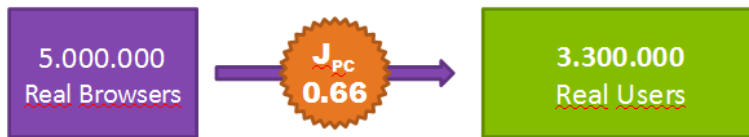
Prístup k internetu z domova a z práce

V rámci OverNight je táto klasifikácia vykonávaná raz týždenne je platná po dobu jedného týždňa. V rámci mesačnej produkcie sa tak môže stať, že jedna IP adresa je jeden týždeň klasifikovaná ako Home a ďalší týždeň ako Work.

7.2 Klasifikácia panelistov

V ďalšom kroku sú panelisti klasifikovaní ako Home alebo Work. Táto klasifikácia je vykonávaná na základe kategórií IP adries a toto priradenie je validované na základe údajov z dotazníka. Najmä sa k tomu využíva otázka, z akých miest sa panelista pripájal k internetu za posledný mesiac z práve používaného zariadenia. Iba ak súhlasí IP klasifikácia s odpoveďou v dotazníku, rozhodne sa o tom, či bude panelista zaradený do platformy Home alebo Work. Panelista môže byť na základe svojej aktivity priradený do oboch platforiem, ak tomu zodpovedá IP klasifikácia a odpovede v dotazníku. Avšak jeho aktivita je vždy započítaná do jednej z platforiem, podľa klasifikácie IP adresy.

PC platforma bez rozdění H/W



Rozdění Home/Work



Vysvetlenie vplyvu rozdelenia Home a Work na RU PC

7.3 Behaviorálna syntéza (BPS) pre Home a Work

Cieľom Behaviorálnej syntézy panelov je všeobecne z jedноплатформových panelov získať multiplatformový panel. Pri tom sa využívajú znalosti z kalibračného panelu, z ktorého sa získavajú vzorce pozitívnych duplikácií, vzorce negatívnej exkluzivity a tiež sa usudzuje na dôležitosť jednotlivých behaviorálnych charakteristík. Opis BPS je v inej časti tohto dokumentu, ale v prípade Home and Work platforiem sú hlavné kroky nasledovné:

1. Pred samotnou syntézou sú obe platformy vážené na štrukturálne dáta, frekvenciu používania internetu, behaviorálne informácie (návštevnosť na zapojených weboch) a počet internetových užívateľov pre každú platformu (informácie získané zo štrukturálneho výskumu).
2. Kalibračný panel sa skladá z panelistov, ktorí sú kategorizovaní v oboch platformách Home a Work.
3. Obe platformy Home a Work sú klastrované na základe pohlavia a veku. Následne v rámci klastrov sú panelisti z oboch platforiem prepojení a spojení za pomoci BPS algoritmu.

Výsledkom je, že platformy Home a Work sú spojené do platformy PC. Táto platforma obsahuje tri typy panelistov: Iba Home, Home a Work, iba Work.

8 Meranie stráveného času (Sonar)

Výpočet stráveného času (metódou Sonar) je riešenie, ktoré sa venuje presnejšiemu vyhodnoteniu času stráveného používateľmi internetu pri prehliadaní obsahu vydavateľov a využívaní ich služieb.

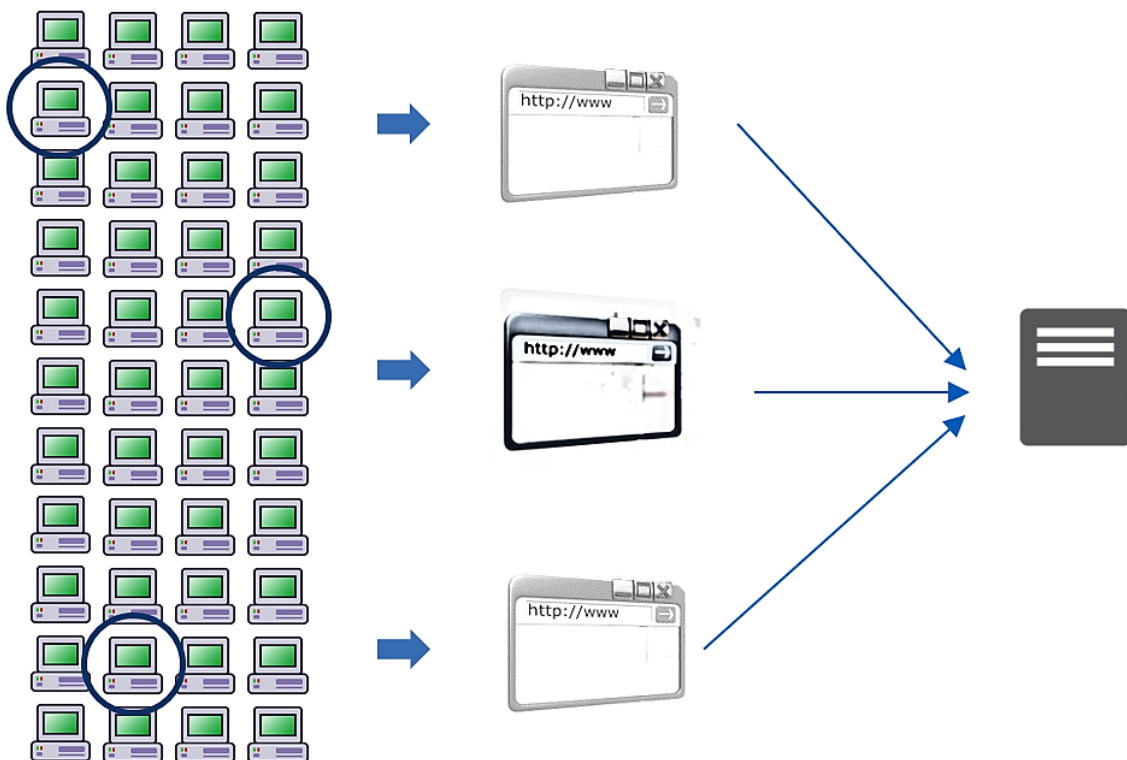
Modul, ktorý používame, rozširuje aktuálne meranie stráveného času publika tak, že pridáva nový typ udalostí: Sonar. Sonary sú vysielané každú sekundu s pravdepodobnosťou 1/3000. Ak sa skript z akéhokoľvek dôvodu neaktívuje, pravdepodobnosť sa odpovedajúcim spôsobom zmení.

Nový prístup znamená, že vypočítame čas strávený na základe vzorky dát. Vzhľadom na to, že sa na webovej stránke vyskytuje veľa zobrazení stránok súčasne, vzorka sonarov stále poskytuje dostatok údajov a súčasne nezaťažuje internetové pripojenie užívateľa, takže odosielanie týchto udalostí je pre používateľov internetu nepozorovateľné.

Zozbierané informácie sa posielajú vo formáte využívajúcom už existujúcu funkciu hlavného skriptu - extraparametre.

Na základe týchto informácií môžeme ľahko vypočítať celkový aktívny čas pre webovú stránku / službu, bez toho, aby sme boli poukazovali na trvanie každého zobrazenia stránky pre každý súbor cookie / BrowserID. Pri Gemius New Time measurement nestratíme informácie o zobrazeniach "non-BrowserID" stránok a ani o posledných zobrazeniach počas návštevy, tak ako pri riešení, ktoré spoločnosť Gemius a tiež aj iné spoločnosti doteraz používali na meranie času.

Obrázok 21: Schéma výpočtu stráveného času



REALIZÁTOR:

Výskumná agentura

Gemius Slovakia, s. r. o.

Kutuzovova 5/A

831 03, Bratislava 3

IABmonitor@gemius.sk

www.gemius.com

OBCHODNÉ ZASTÚPENIE:

Združenie pre internetovú reklamu

IAB Slovakia

Zochova 754 / 6-8

811 03, Bratislava 1

iabmonitor@iabslovakia.sk

www.iabslovakia.com