

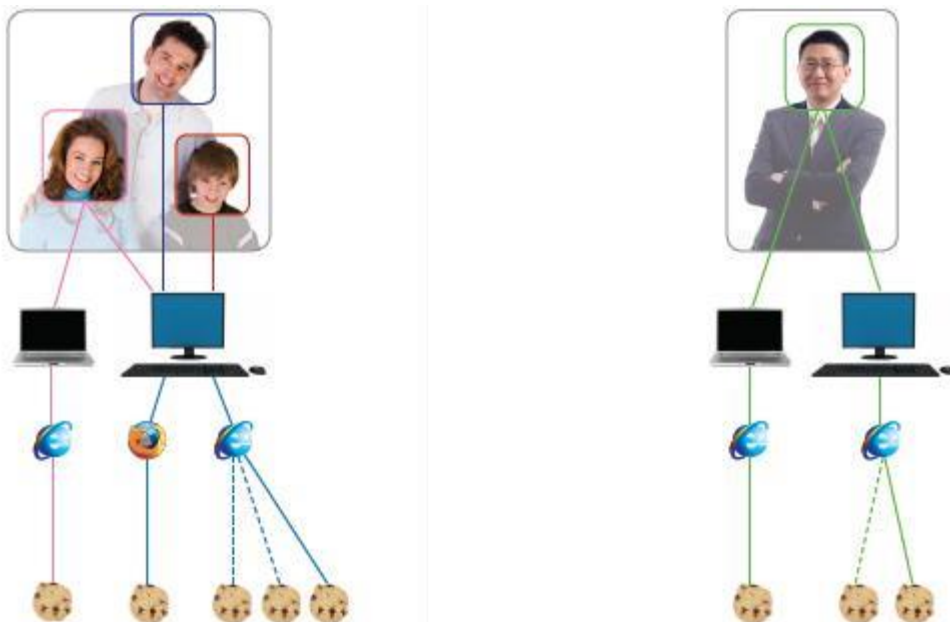
GemiusAudience metodologija - Dobivanje brojki jedinstvenih korisnika (Real Users)

*Gemius d.o.o.
Ulica grada Vukovara 254
10000 Zagreb - Hrvatska*

Za razumijevanje metodologije dobivanja real usera potrebno je određeno tehničko znanje, kao npr. što su to cookie-i, analitički (iliti site-centric) alati i slično.

PROBLEM: U Internet svijetu postoje idući "fenomeni": brisanje cookiea, dijeljenje istog korisničkog računa na računalo (jedan cookie je dodjeljen za više osoba koje koriste isto računalo), korištenje više računala/pametnih telefona/Internet preglednika/ (web browser-a) od strane jedne osobe (više od jednog cookie-a je dodjeljeno jednoj te istoj osobi).

PRIMJERI:



- 1 računalo kojeg koristi 1 osoba
- 1 računalo kojeg koriste 3 osobe
- 3 preglednika generiraju 5 cookie-a u u datom mjesecu datom mjesecu
- 2 računala koja koristi 1 osoba
- 2 preglednika generiraju 3 cookie-a

EFEKT: Broj cookie-a nikako **ne može** biti dobar pokazatelj broja jedinstvenih Internet korisnika. Što se duži vremenski period analizira, primjećuje se veća razlika u brojkama između cookie-a i jedinstvenih Internet korisnika. Broj svih cookie-a na nekom tržištu (državi) tokom mjeseca može biti i do 10 puta veći nego broj jedinstvenih Internet korisnika na tom tržištu.

Na primjer, u Hrvatskoj broj tuzemnih cookie-a na mjesečnoj razini iznosi oko 17 miliona (podaci iz 01/2015), a u Bosni i Hercegovini broj tuzemni cookie-a na mjesečnoj razini iznosi oko 7 miliona (podaci iz 01/2015).

RIJEŠENJE: Gemius koristi algoritam "jedinstvenih korisnika" (Real Users) za izračun broja individualnih osoba koje posjećuju Internet stranice.

Metrika "jedinstveni korisnici" odnosi se na broj osoba koji su posjetili datu Internet stranicu u analiziranom vremenskom razdoblju bez obzira na učestalost (frequency) posjete toj Internet stranici, broj uređaja odnosno cookie-a sa kojima su posjetili tu Internet stranicu te koliko vremena su proveli (time spent) na toj Internet stranici u analiziranom vremenskom razdoblju.

JEDINSTVENI POSJETITELJI - ALGORITAM

1. Filtriranje podataka orijentiranih prema Internet stranici (podaci prikupljeni sa gemiusTraffic analitičkim alatom)

Site-centric iliti analitički alat skuplja informacije o prometu svih Internet stranica na koje su postavljene gemiusTraffic skripte (Gemiusov analitički alat, kao npr. googleAnalytics od Google-a). Ovi podaci se filtriraju za daljnju uporabu za gemiusAudience istraživanje. Kako bi dobili podatke važeće za gemiusAudience istraživanje provode se idući koraci:

- uklonjene su sve posjete iz inozemstva (bazirane na IP adresama)
- nedozvoljeni iframe promet je uklonjen
- samoosvježavajući (autorefresh) promet je uklonjen
- promet na ne dozvoljenim domenama je uklonjen

Svi daljnji koraci u gemiusAudience istraživanju provedeni su na prometu koji je ovako filtriran!

2. Izračun SUA (Speed Up Algorithm) vrijednosti

SUA je jednostavno naziv našeg algoritma, a vrijednost istoga je izračunata:

- zasebno za svaku Internet stranicu koja sudjeluje u gemiusAudience istraživanju (**SUAwebsite**);
- ukupno za sve Internet stranice koje sudjeluju u gemiusAudience istraživanju (**SUAtotal**).

SUA vrijednost stoji za minimalan broj kolačića (cookie-a) za koje je sigurno da su postojali kako bi generirali sve prikaze stranica (page views) svih Internet stranica koje sudjeluju u istraživanju (uključujući u obzir činjenicu o brisanju cookie-a te da je jedan obrisan cookie zamijenjen novim).

Kako bi prebrojali te cookie, idući koraci su provedeni:

- a. Za svaki domaći (tuzemni) cookie, vrijeme njegovog stvaranja (creation timestamp) i zadnji "PV" (Page View) su provjereni (vrijeme PV-a znači vrijeme zadnjeg pregleda stranice kojeg je učinio cookie u mjerenom mjesecu na bilo kojoj Internet stranici

koja sudjeluje u istraživanju). Vrijeme stvaranja cookie-a se razlikuje od zadnjeg "PV" vremena.

- b. Kako bi brojali *SUAtotal* vrijednost svi creation timestampovi na tržištu su odabrani, a kako bi brojali *SUAwebsite* vrijednost odabrani su samo oni creation timestampovi dodjeljeni cookieima koji su bili aktivni na mjerenoj Internet stranici u mjerenom mjesecu.
- c. Odabrani creation timestampovi sortirani su uzlazno
- d. Početna vrijednost X-a u algoritmu je 0
- e. Svi timestampovi su provjereni od početka: ukoliko se pojavi novi cookie (provjereni timestamp je vrijeme stvaranja) vrijednost X-a je povećana za 1. Ukoliko se pojavi timestamp zadnjeg "PV-a" vrijednost X-a je umanjena za 1.
- f. Zadnja vrijednost X-a u algoritmu je 0
- g. SUA vrijednost jednaka je maksimumu vrijednosti X-a

3. Izračun dosega (REACH) svake Internet stranice

Doseg Internet stranice računa se formulom:

$$REACH_{website} = SUA_{website} / SUA_{total} * REACH_{total}$$

Vrijednosti *SUAwebsite* i *SUAtotal* su poznate (iz točke 2.).

REACHtotal je doseg gemiusAudience istraživanja na tržištu u mjerenom mjesecu, što znači postotak Internet korisnika koji posjećuju barem jednu od Internet stranica uključenih u gemiusAudience istraživanje barem jedanput u mjerenom mjesecu. **REACHtotal** je jedina brojka koju dobivamo procjenom i u svrhu primjera definirat ćemo da je reach gemiusAudience istraživanja na datom tržištu 90%.

4. Završno brojanje vrijednosti jedinstvenih korisnika (Real usera)

Veličina populacije Internet korisnika na tržištu (**POPULATION**) u mjerenom mjesecu poznata je od prije (**dobiva se strukturalnom**

studijom) i u svrhu primjera definirat ćemo da je Population 10 000 000.

Ukupan broj jedinstvenih korisnika u mjenom mjesecu za gemiusAudience istraživanje računa se formulom:

$$RU_{total} = REACH_{total} * POPULATION$$

Broj jedinstvenih korisnika (Real Users) za neku Internet stranicu za mjereni mjesec računa se formulom:

$$RU_{website} = REACH_{website} * POPULATION$$

PRIMJER

1. Filtriranje podataka orijentiranih prema Internet stranici (podaci prikupljeni sa gemiusTraffic analitičkim alatom)

Filtriranje se sastoji od nekoliko koraka.

Svi cookie-i su podjeljeni na domaće i inozemne temeljem IP adresa. Također, svi non-cookie-i pregledi stranica (page views) su podjeljeni na domaće i inozemne. Samo je domaći promet prihvaćen u gemiusAudience istraživanju.

Većina Internet stranica koje sudjeluju u istraživanju ima stvorenu svoju definiciju (struktura u gemiusTraffic-u tj. kategorizacija): dozvoljene domene i dozvoljeni iframe-ovi (ukoliko postoje) su definirani – promet koji ne ispunjava ove kriterije je isključen iz gemiusAudience sirovih podataka (raw data).

Samoosvježavajući promet je također otkriven i uklonjen iz gemiusAudience sirovih podataka.

Zbog procesa filtriranja broj otvorenih stranica (page views) i drugih metrika (Visits, Time, itd.) u gemiusAudience istraživanju može biti manji, i obično je, nego broj otvorenih stranica zabilježen

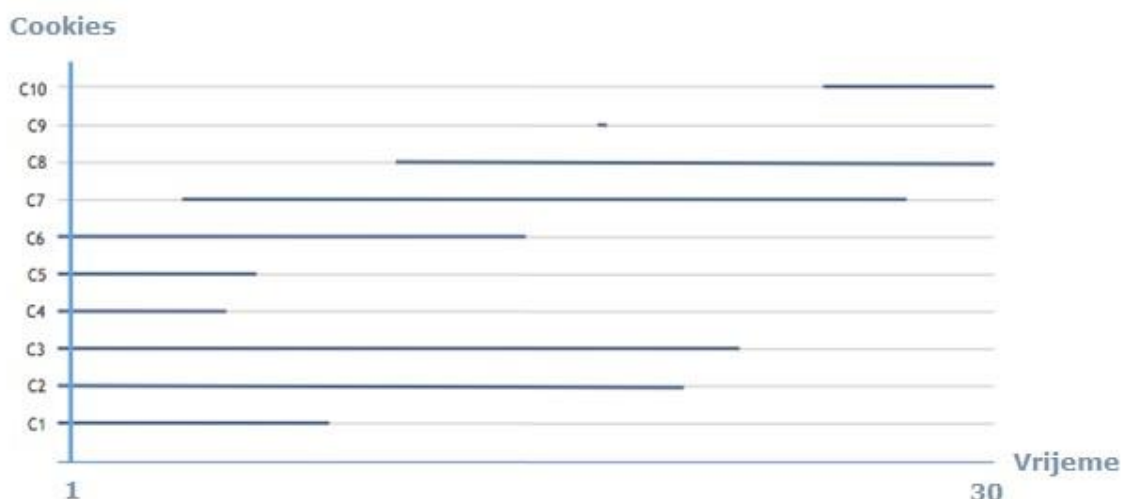
od strane site-centric alata (npr. gemiusTraffic-a ili googleAnalytics-a).

2. Izračun SUA vrijednosti

Poslije procesa filtriranja cijelog prometa (opisano u koraku iznad), na tržištu se stvorilo npr. 10 cookie-a: C1, C2, C3, C4, C5, C6, C7, C8, C9 i C10.

Svaki od ovih 10 cookie-a ima definirana dva vremena stvaranja (iliti dva timestamp-a):

- **Vrijeme stvaranja (creation timestamp)** – Ukoliko je cookie stvoren prije mjenog mjeseca prvo je definiran timestamp mjenog mjeseca. U tablici ispod cookie-i C1, C2, C3, C4, C5 i C6 su stvoreni prije mjenog mjeseca pa je njihov timestamp postavljen kao početak mjeseca (kao prva sekunda u mjesecu). Preostali cookie-i su stvoreni tokom mjenog mjeseca.
- **Zadnji "Page View" timestamp** – Timestamp zadnjeg pregleda stranice zadanog cookie-a u mjenom mjesecu.



Treba spomenuti da je zadan cookie mogao posjetiti druge Internet stranice tokom mjeseca, ali njegov creation timestamp i zadnji "PV" timestamp su definirani samo jedanput ito analizom prometa svih Internet stranica koje sudjeluju u gemiusAudience istraživanju na zadanom tržištu.

To znači da je na primjer cookie "C1" mogao posjetiti Internet stranicu "W1" samo jedanput i napraviti samo jedan prikaz stranice (page view) na toj Internet stranici, ali njegov creation timestamp i zadnji "PV" timestamp su definirani analizom njegovog prometa na svim Internet stranicama u gemiusAudience istraživanju, pa je njegov prvi prikaz stranice (onaj koji se koristi za definiranje creation timestamp-a) mogao biti stvoren na Internet stranici "W2", a njegov zadnji prikaz stranice u mjesecu je mogao biti stvoren na Internet stranici "W3".



SUA vrijednost izračunata je za cijelo Internet tržište:

- Početna vrijednost X-a u algoritmu je 0;
- Svi timestampovi su provjereni od početka: Ukoliko se pojavi novi cookie (provjereno vrijeme timestampa je vrijeme stvaranja) vrijednost X-a povećana je za jedan (+1), a ukoliko je timestamp zadnji "PV" timestamp vrijednost X-a smanjena je za jedan (-1);
- Zadnja vrijednost X-a u algoritmu je 0

Tablica ispod predstavlja "+1" za cookie sa creation timestampom i "-1" za cookie sa zadnjim "PV" timestampom.



Tablica ispod prikazuje računanje vrijednosti X-a temeljem tablice iznad **(vrijeme se gleda kronološki, od prve sekunde u mjesecu, do zadnje)**

	0	+1 C1	+1 C2	+1 C3	+1 C4	+1 C5	+1 C6	+1 C7	-1 C4	-1 C5	-1 C1	+1 C8	-1 C6	+1 C9	-1 C9	-1 C2	-1 C3	+1 C10	-1 C7	-1 C8	-1 C10
X	0	1	2	3	4	5	6	7	6	5	4	5	4	5	4	3	2	3	2	1	0

Maksimalna vrijednost X-a za cijelo Internet tržište je 7. Dakle ***SUA_{total}*=7.**

U idućem koraku izračunat ćemo SUA vrijednost za pojedinu Internet stranicu (***SUA_{website}***). Ovdje je metodologija za računanje istao kao i za cijelo Internet tržište odnosno ***SUA_{total}*** (objašnjeno iznad). Imajte na umu kako je broj cookie-a na nekoj Internet stranici uvijek manji nego broj cookie-a na datom Internet tržištu. Drugim riječima ***SUA_{website}*** vrijednost je uvijek manja od ***SUA_{total}*** vrijednosti.

Zamislimo da su Internet stranicu W1 posjetili cookie-i C1, C3, C6, C7 i C10. Tablica ispod predstavlja podatke samo za W1 Internet stranicu. Imajte na umu da su creation timestamp i zadnji "PV" timestamp za cookie zajednički za cijelo tržište (i sve Internet stranice u

gemiusAudience istraživanju sudjeluju jednako u procesu definiranja ovih timestampova).



	0	+1 C1	+1 C3	+1 C6	+1 C7	-1 C1	-1 C6	-1 C3	+1 C10	-1 C7	-1 C10
X	0	1	2	3	4	3	2	1	2	1	0

Maksimalna vrijednost X-a za Internet stranicu W1 je 4,. Dakle **SUAwebsite1=4**.

3. Izračun doseg svake Internet stranice

Doseg gemiusAudience istraživanja u analiziranom mjesecu:

$$REACHtotal = 90\% \quad SUAwebsite = 4 \quad SUAtotal = 7$$

Doseg Internet stranice W1 iznosi:

$$REACHwebsite1 = 4/7 * 90\% = 51,42\%$$

4. Završno brojanje vrijednosti jedinstvenih korisnika

Veličina populacije Internet korisnika u analiziranom mjesecu za dato držanje:

$$POPULATION = 10\ 000\ 000$$

Ukupan broj jedinstvenih korisnika (Real Users) u analiziranom mjesecu:

$$RUtotal = 90\% * 10\ 000\ 000 = 9\ 000\ 000$$

Broj jedinstvenih korisnika za Internet stranicu W1 u analiziranom mjesecu:

$$RUwebsite1 = 51,42\% * 10\ 000\ 000 = 5\ 142\ 000$$

NE ZABORAVITE

- U gemiusAudience istraživanju sudjeluje samo domaći (tuzemni) promet koji je filtriran od nedozvoljenog prometa.
- Populacija (POPULATION) tj. ukupan broj Internet korisnika na tržištu dobiva se iz offline strukturalnih studija (terenska istraživanja – CATI/CAPI).
- REACHtotal je jedina brojka koju dobivamo procjenom.
- Ne zaboravite, formule su bitne:
$$\text{REACHwebsite} = \text{SUAtotal} / \text{SUAtotal} * \text{REACHtotal}$$
$$\text{RUwebsite} = \text{REACHwebsite} * \text{POPULATION}$$
$$\text{RUtotal} = \text{REACHtotal} * \text{POPULATION}$$
- Svi analitički alati, a ovdje fokus stavljamo na googleAnalytics jer je najpopularniji, mjere cookie i prezentiraju cookie kao jedinstvene korisnike, a što nikako ne može biti ispravno. Dakle metrika "Unique visitors", odnosno po novome "Users", bazirana je na cookie-ima
(<https://developers.google.com/analytics/resources/concepts/gaConceptsCookies>), a znamo da jedan stvarni (jedinstveni) korisnik može generirati, i generira, više cookie-a, npr. 3. Prema tome googleAnalytics daje brojku od 3 stvarna korisnika (Unique visitor-a iliti User-a), dok gemiusAudience istraživanje postoji kako bi dobili ispravne podatke o jedinstvenim korisnicima.

KONTAKT

Gemius d.o.o.

Ulica grada Vukovara 254

10000 Zagreb – Hrvatska

Tel: +385 1 4667 321

Vibor Kalogjera: vibor.kalogjera@gemius.com, +385 99 218 9601

Josip Gorički: josip.goricki@gemius.com, +385 98 175 4619