

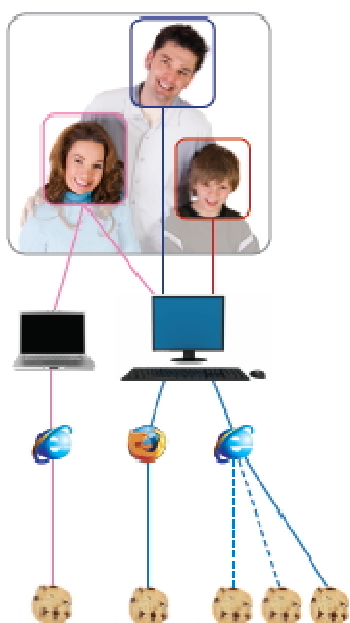
Dobivanje brojki jedinstvenih korisnika u gemiusAudience istraživanju

Marinko Penava, Gemius d.o.o.

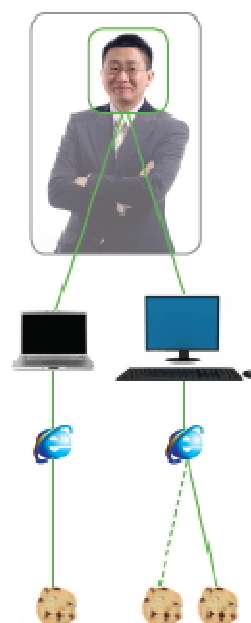
Ulica grada Vukovara 254, 10000 Zagreb

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

PRIMJERI:



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



- 2 računala koja koristi 1 osoba
- 2 preglednika generiraju 3 cookie-a u datom mjesecu

EFEKT: Broj cookie-a **nije** 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.

RIJEŠENJE: Gemius koristi algoritam "jedinstvenih korisnika" za izračun broja individualnih osoba koje posjećuju Internet stranice. Metrika "jedinstveni korisnici" odnosi se na broj osoba koji su posjetili datu Internet stranicu barem u analiziranom vremenskom razdoblju bez obzira na učestalost (frequency) posjete toj Internet stranici i koliko vremena su proveli (time spent) na toj Internet stranici u analiziranom vremenskom razdoblju.

JEDINSTVENI POSJETITELJI - ALGORITAM

1. Filtriranje podataka orjenitanih prema Internet stranici (Filtering site-centric data – analitički alat)

Site-centric (analitički alat) skuplja informacije o prometu svih Internet stranica na koje su postavljene gemiusTraffic skripte. Ovi podaci su filtrirani 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 ovako filtriranom prometu.

2. Izračun LBU vrijednosti

LBU vrijednost je izračunata:

- zasebno za svaku Internet stranicu koja sudjeluje u gemiusAudience istraživanju (**LBUwebsite**);
- Skupa za sve Internet stranice koje sudjeluju u gemiusAudience istraživanju (**LBUtotal**).

LBU vrijednost stoji za minimalan broj kolačića (cookie) za koje je sigurno da su postojali kako bi generirali sve prikaze stranica (page views) svih Internet stanica koje sudjeluju u istraživanju

(uključujući u obzir činjenicu o brisanju cookie-a i da je jedan obrisani cookie zamijenjen novim).

LBU vrijednost jednaka je maksimalnom broju cookie-a koji su postojali u isto vrijeme (u jednoj sekundi) u mjerenom mjesecu. 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 *LBUtotal* svi creation timestampovi su odabrani, a kako bi brojali *LBUwebsite* odabrani su samo oni creation timestampovi dodjeljeni cookie-ima koji su bili aktivni na mjerenoj Internet stranici u mjerenom mjesecu.
- c. Odabrani creation timestampovi sortirani su uzlazno
- d. Početna algoritam vrijednost X-a 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. LBU vrijednost je jednaka maksimumu vrijednosti X-a

3. Izračun dosega (reach) svake Internet stranice

Doseg Internet stranice jednak je:

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

Vrijednosti *LBUwebsite* i *LBUtotal* 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.

4. Završno brojanje vrijednosti jedinstvenih korisnika

Veličina populacije Internet korisnika na tržištu (**POPULATION**) u mjerenom mjesecu poznata je od prije (**iz offline strukturalnih studija za zadano tržište**).

Ukupan broj jedinstvenih korisnika u mjerenom mjesecu za gemiusAudience istraživanje poznat je:

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

Broj jedinstvenih korisnika za neku Internet stranicu za mjereni mjesec izračunava se:

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

PRIMJER

1. Filtriranje podataka orjenitiranih prema Internet stranici (Filtering site-centric data - analitički alat)

Filtriranje se sastoji od nekoliko akcija.

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!

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

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

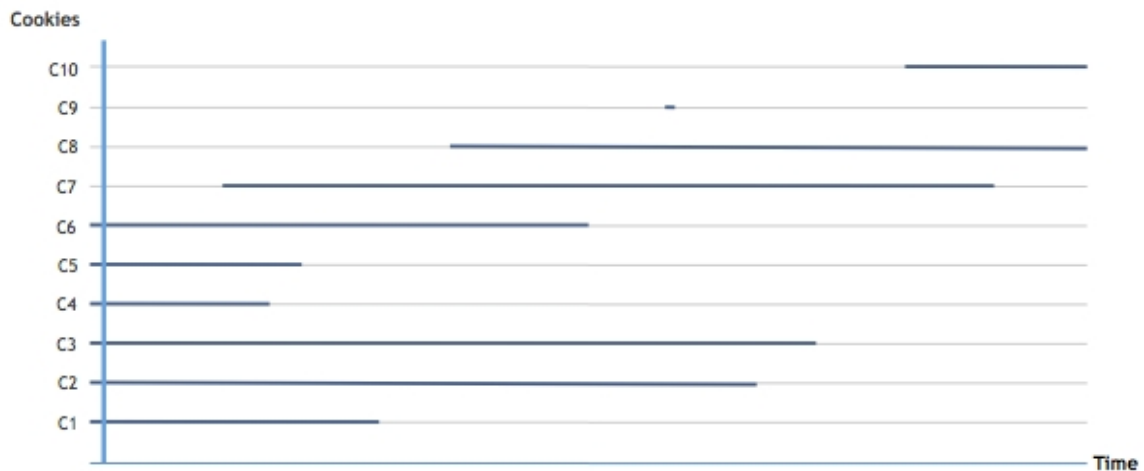
Zbog procesa filtriranja broj otvorenih stranica u gemiusAudience istraživanju može biti manji (i najčešće je) nego broj otvorenih stranica zabilježen od strane site-centric alata (analitičkog alata - gemiusTraffic).

2. Izračun LBU vrijednosti

Poslije procesa filtriranja cijelog prometa (opisano iznad u koraku 1.), na tržištu se stvorilo 10 cookie-a: C1, C2, C3, ..., C10.

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

- vrijeme stvaranja – ukoliko je cookie stvoren prije mjerenog mjeseca prvo je definiran timestamp mjerenog mjeseca. U tablici ispod cookie-i C1, C2, C3, C4, C5 i C6 su stvoreni prije mjerenog mjeseca pa je njihov timestamp postavljen kao početak mjeseca. Preostali cookie-i su stvoreni tokom mjerenog mjeseca.
- zadnji "PV" (last PV) timestamp – timestamp zadnjeg pregleda stranice zadanog cookie-a u mjerenom mjesecu.



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

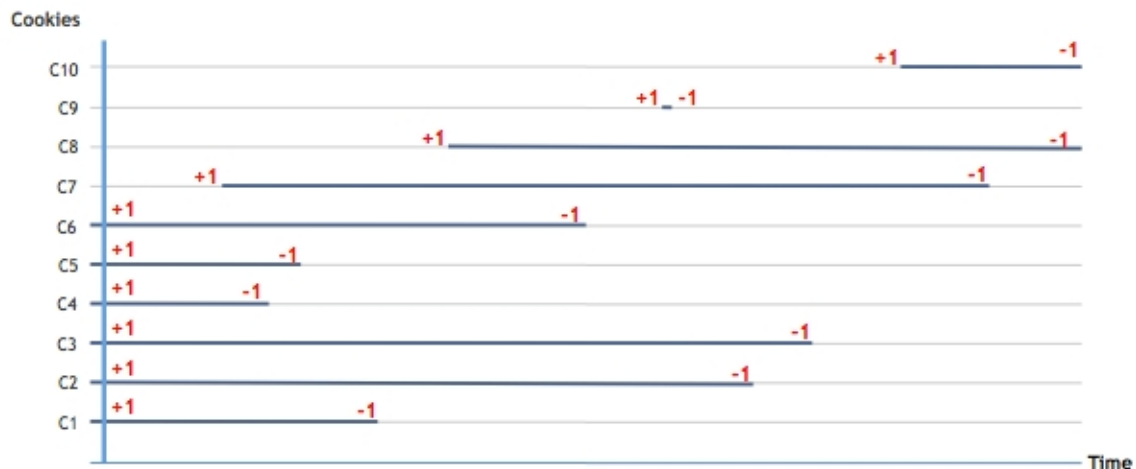
To znači da je na primjer cookie "C1" mogao posjetiti Internet stranicu "W1" samo jedanput i napraviti samo jedan prikaz stranice (page view) na toj Internet stranici, ali njegov creation timestamp i zadnji "PV" timestamp su definirani analizom njegovog prometa na svim Internet stranicama u istraživanju, 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".



LBU vrijednost je izračunata za cijeli Internet:

- početna vrijednost algoritma X 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 1, a ukoliko je timestamp zadnji "PV" timestamp vrijednost X-a smanjena je za 1;
- zadnja vrijednost X-a u algoritmu je 0

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

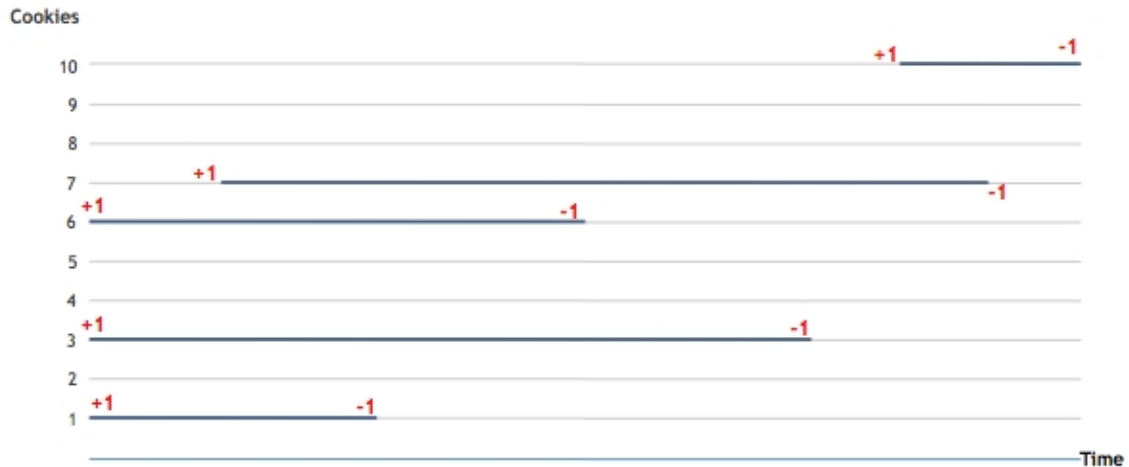


Tablica ispod predstavlja računanje vrijednosti X.

	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 $LBU_{total}=7$.

U idućem koraku LBU vrijednost za pojedine Internet stranice je stvorena. Zamislamo da su Internet stranicu W1 posjetili cookie-i C1, C3, C6, C7 I C10. Tablica ispod predstavlja podatke samo za te cookie. Imajte na umu da su creation timestamp i zadnji "PV" timestamp za cookie zajednički za cijelo tržište (i sve Internet stranice 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, pa je $LBU_{website1}=4$.

3. Izračun doseg svake Internet stranice

Doseg gemiusAudience istraživanja u analiziranom mjesecu:
 $REACH_{total} = 90\%$

$$LBU_{website} = 4$$

$$LBU_{total} = 7$$

Doseg Internet stranice W1 iznosi:

$$REACH_{website1} = 4/7 * 90\% = 51,42\%$$

4. Završno brojanje vrijednosti jedinstvenih korisnika

Veličina populacije Internet korisnika u analiziranom mjesecu za dato držište:

$$POPULATION = 10\,000\,000$$

Ukupan broj jedinstvenih korisnika u analiziranom mjesecu:

$$RU_{total} = 90\% * 10\,000\,000 = 9\,000\,000$$

Broj jedinstvenih korisnika za Internet stranicu W1 u analiziranom mjesecu:

$$RU_{website1} = 51,42\% * 10\,000\,000 = 5\,142\,000$$

GOOD TO KNOW

- U gemiusAudience istraživanju sudjeluje samo domaći (tuzemni) promet koji je filtriran od nedozvoljenog iframe prometa i samoosvježavajućeg prometa.
- Populacija (POPULATION) tj. ukupan broj Internet korisnika na tržištu dobiva se iz offline strukturalnih studija (terenska istraživanja).
- Google Analytics "Unique visitors" baziran je na cookie-ima (<https://developers.google.com/analytics/resources/concepts/gaConceptsCookies>), a znamo da jedan stvarni (jedinstveni) korisnik može generirati npr. 3 cookie-a. Prema tome Google Analytics daje brojku od 3 stvarna korisnika (unique visitora), dok gemiusAudience daje brojku od 1 stvarnog korisnika.

Google Analytics Unique Users \neq GemiusAudience Real Users