

WPROWADZENIE DO BADANIA MEDIAPANEL

1.	Podstawowe informacje o Badaniu	2
2.	Panele w Badaniu	4
3.	Populacja podlegająca Badaniu	5
4.	Główne elementy metodologii badania	5
4.1.	Constant Panel	5
4.2.	Metodologia liczenia czasu w Internecie (Sonary).....	6
4.3.	Współczynnik J BRUS - średnia liczba przeglądarek na użytkownika.....	7
4.4.	Zasięg audytowanej części badania i sposób wyznaczania populacji (Metoda PRES)	7
4.5.	Odwzorowanie liczby odsłon, czasu i wizyt dla witryn i aplikacji audytowanych (Algorytm FPA 2.0)	8
4.6.	Sposób estymacji liczby przeglądarek (EC Global)	9
4.7.	Metodologia produkcji wyników dla witryn niepodlegających audytowi site-centric	10
4.	Dane w interfejsie mediapanel.gemius.com.....	11
5.	Publikacja danych.....	11

1. Podstawowe informacje o Badaniu

Celem badania jednoźródłowego Mediapanel jest dostarczenie informacji o korzystaniu z Internetu, słuchalności Radia oraz oglądalności Telewizji w Polsce.

1) **Pomiar Internetu** odbywa się poprzez połączenie pomiarów user-centric i site-centric:

a. User-centric

Zbiera informacje o zachowaniu internautów biorących udział w badaniu, tzw. panelistów (nt. odwiedzania witryn internetowych i korzystania z aplikacji). Dane na temat aktywności panelistów są zbierane za pomocą specjalnego oprogramowania raportującego NetPanel (software panel), zaś na witrynach audytowanych dodatkowo przy udziale skryptów zliczających (BID panel). Informacje z obu typów paneli łączone są w procesie fuzji, w wyniku której powstaje wspólny panel fuzyjny.

b. BPS – Behavioral Panel Synthesis

Panele na poszczególnych platformach mają część wspólną – panel kalibracyjny. To grupa osób, u których pomiarem objęte są wszystkie używane przez nich urządzenia. Na podstawie zachowania tych osób algorytm BPS jest w stanie z dużym prawdopodobieństwem wskazać tych samych użytkowników na różnych urządzeniach. Dzięki temu możliwe jest pokazanie wspólnego zasięgu poszczególnych witryn na wszystkich platformach.

c. Site-centric (gemiusPrism)

Opiera się na technologii BID, pozwalającej na identyfikowanie użytkowników monitorowanych witryn bez naruszania ich prywatności. Dane uzyskiwane są za pomocą skryptów zliczających umieszczonych w kodzie HTML witryn audytowanych, czyli włączonych do badania z wykorzystaniem danych z tych skryptów. Informacje o liczbie użytkowników oraz dokonywanych przez nich odsłonach przesyłane są automatycznie do centrum obliczeniowego Gemius, gdzie następuje bieżąca aktualizacja wyników.

Ponieważ badanie gemiusPrism (zbieranie informacji o ruchu na witrynach) jest pomiarem obejmującym pełną aktywność użytkowników audytowanych witryn (ma charakter spisowy), jest to metoda pozbawiona typowych błędów statystycznych badań na próbach. Dzięki temu wyniki badania internetowego dla witryn poddanych audytowi site-centric są maksymalnie zbliżone do rzeczywistości pod względem liczby odsłon, wizyt i spędzonego na witrynach czasu.

2) Pomiar Radia i Telewizji

Opiera się na pasywnym pomiarze zachowań respondentów z panelu badawczego.

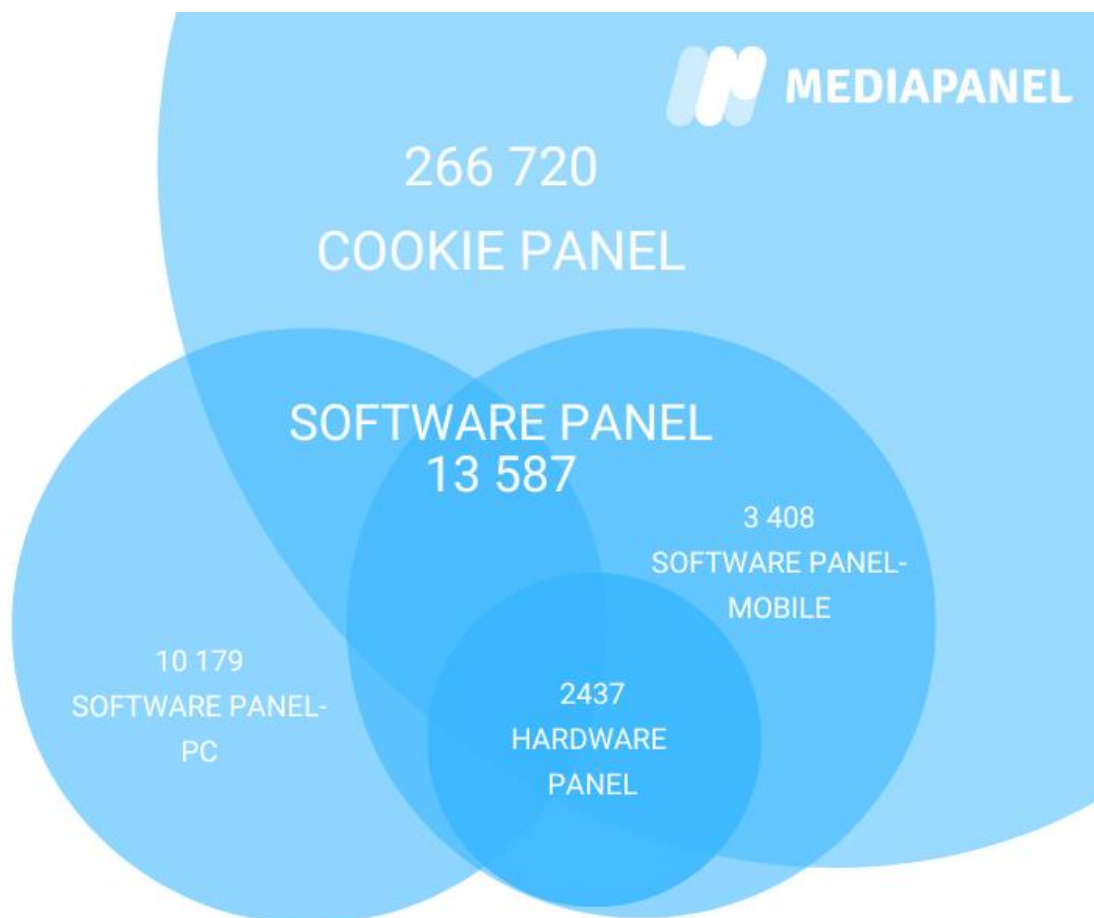
Wykorzystujemy audiomatching, czyli dopasowanie sygnatur dźwięku otoczenia panelisty do sygnatur zebranych w Centrum Sygnatur Referencyjnych Gemius. Dopasowanie kanałów TV oraz programów radiowych działa w modelu 24/7. Sygnatury audio zbierane są z częstotliwością 30 na sekundę, w przypadku braku połączenia miernika z Internetem są buforowane i wysyłane później (do 7 dni).

Miernikiem jest smartfon z dedykowanym oprogramowaniem Gemius, instalowanym na poziomie systemu operacyjnego, który umożliwia m.in.:

- generowanie sygnatur audio wykorzystywanych w audiomatchingu,
- pomiar lokalizacji miernika poprzez koordynat GPS (jako podstawa do wyznaczania oglądalności poza domem),
- pomiar uruchomionych aplikacji, w tym przeglądarek,
- monitorowanie korzystania ze słuchawek,
- pomiar naładowania baterii, zmiany karty SIM.

Panel ten nazywany jest Hardware Panelem.

2. Panele w Badaniu



- 1) **Panel Cookie (BID)** – dane zbierane za pomocą kwestionariuszy Gemius wyświetlanych na oskryptowanych stronach www (urządzenia PC, Phones, Tablets).

Paneliści Cookie nie ukończyli procesu rejestracji i nie zainstalowali oprogramowania NetPanel, lecz pozostawili nam swoje dane socjo-demograficzne.

- 2) **Software Panel** – dane zbierane za pomocą dedykowanej aplikacji na platformach PC, Phones, Tablets, tzw. oprogramowanie NetPanel.
- 3) **Panel Hardware** – dane zbierane za pomocą miernika na poziomie systemu operacyjnego smartfonu.

Smartfony z miernikiem służącym do audiomatchingu monitorują także aktywność panelistów (odwiedzanie witryn internetowych i korzystanie z aplikacji), tym samym powiększają Software Panel – mobile.

Osoby będące Hardware Panelistami proszone są także o zainstalowanie oprogramowania NetPanel na swoim komputerze i tym samym powiększają Software Panel – PC.

3. Populacja podlegająca Badaniu

Populacja podlegająca badaniu to „Polacy w wieku 7 - 75”.

Aktualną podstawą do wyznaczania populacji internautów w badaniu Mediapanel są najświeższe dane GUS oraz dane z Badania NetTrack.

Behawioralnie panel reprezentatywny jest dla tej części populacji, która deklaruje siebie jako użytkowników Internetu i/lub posiadaczy smartfonów/telefonów komórkowych.

4. Główne elementy metodologii badania

Produkcja wyników odbywa się codziennie (za dzień wczorajszy).

Dane z Hardware Panelu podlegają walidacji po stronie danych demograficznych oraz danych behawioralnych – są weryfikowane pod względem poprawności działania miernika w danym dniu. Po walidacji dane te tworzą panel netto, który podlega procesowi ważenia. Na tym etapie dzienne wyniki dla platform TV oraz Radio łączone są z wynikami dla platformy Internet. W celu zaprezentowania danych dla wszystkich mediów (Internet, Radio, TV) w dowolnych okresach metodologia badania Internetu została rozbudowana o metodę Constant Panel.

4.1. Constant Panel

Constant Panel to wirtualny panel, który zawiera bardzo dużą liczbę panelistów o równej wadze. Zamiast zmieniać wagę jednego panelisty, symulujemy większą liczbę panelistów z dużych grup celowych i mniejszą liczbę panelistów z małych grup celowych.

Panel nazywa się „stałym”, ponieważ waga panelistów nie zmienia się w czasie, a ich liczba podlega modyfikacjom jedynie w przypadku wzrostów populacji lub zmian struktury. Symulujemy zatem bezpośrednio sytuację z populacji w świecie rzeczywistym, gdzie wszyscy ludzie są liczeni jednakowo, a naturalna rotacja jest dość niska. Dzięki tej stabilności możemy łatwo obliczyć spójne wyniki dla dowolnego niestandardowego okresu bezpośrednio z panelu.

Istotą tej metodologii jest przypisanie wirtualnemu paneliście aktywności panelu BPS. W fuzji dziennej, która jest sercem metodologii Constant Panel, dopisywana jest aktywność w taki sposób, żeby jednocześnie kontrolować wartości dzienne i 28 dniowe.

Liczba Constant Panelistów w Polsce wynosi 200 tysięcy.

4.2. Metodologia liczenia czasu w Internecie (Sonary)

W wynikach badania prezentowany jest jedynie czas korzystania z aktywnych zakładek przeglądarki.

Zastosowane rozwiązanie służące do zliczania czasu polega na cyklicznym wywoływaniu eventów na aktywnych zakładkach, które wysyłane są z prawdopodobieństwem 1/40 w każdej sekundzie (są to tzw. „sonary”). Sonary zaimplementowane są w skrypcie xgemius.js , który od wersji 3.8 przesyła hity sonarowe. Pomiar ten odbywa się na stronach www, które posiadają zaimplementowany skrypt audytu site-centric.

Dla witryn i aplikacji nieaudytowanych liczony jest czas aktywny na podstawie danych pochodzących z miernika software’owego.

Gemius planuje wdrożenie modyfikacji SDK, która pozwoli na liczenie czasu aktywnego aplikacji audytowanych. Obecnie czas dla aplikacji audytowanych jest liczony na podstawie wizyt, czyli sumowane są czasy pomiędzy kolejnymi odsłonami wykonanymi w aplikacji. Jeżeli aplikacja zawiera oskryptowane materiały stream, to hity streamowe również biorą udział w liczeniu czasu, aby dobrze uwzględnić pełen czas korzystania z tych materiałów, nawet jeżeli pomiędzy odsłonami minie ponad 30 minut.

Prezentowane dane na temat czasu zbierane w następujący sposób:

<u>Witryny audytowane – platforma PC i mobile</u> czas aktywny na podstawie sonarów	Od 01.08.2020 – aktualnie
<u>Witryny nieaudytowane – platforma PC</u> czas aktywny na podstawie miernika software	Od 01.08.2020 – aktualnie
<u>Witryny nieaudytowane – platforma mobile</u> czas aktywny na podstawie miernika software	Od 1.06.2021 - aktualnie
<u>Aplikacje audytowane</u> czas na podstawie wizyt	Od 1.09.2019 – aktualnie
<u>Aplikacje nieaudytowane</u> czas aktywny na podstawie miernika software	Od 1.06.2021 - aktualnie

4.3. Współczynnik J BRUS - średnia liczba przeglądarek na użytkownika

J BRUS to współczynnik, odpowiadający na pytanie, jaka jest średnia liczba internautów korzystających z jednej przeglądarki w odniesieniu do danego serwisu oraz platformy. Wartość współczynnika zależy od dwóch czynników: specyfiki danej platformy (PC, telefony, tablety) oraz liczby użytkowników analizowanego węzła.

Współczynnik przekłada liczbę przeglądarek na liczbę realnych użytkowników (RU) i wyliczany jest dla każdej witryny i aplikacji oddzielnie.

4.4. Zasięg audytowanej części badania i sposób wyznaczania populacji (Metoda PRES)

Zasięg

Zasięg wyznaczamy sprawdzając, jaka część panelistów odwiedza węzły audytowane. Zakładamy, że Internauci odwiedzają podobne witryny niezależnie od urządzenia i przyjmujemy Platformę PC jako wyznacznik zasięgu innych platform.

Obecnie przyjmujemy, że zasięg wynosi 90%.

W przypadku wystąpienia istotnych zmian wskaźnik ten będzie aktualizowany wraz z danymi za styczeń na podstawie danych z roku poprzedniego.

Populacja

Do wyznaczenia aktywnej populacji w danym Cyklu Produkcyjnym używane są informacje z ważenia danych panelowych. Dane te dowożone są do:

- populacji 28 dniowej audytowanej oraz nieaudytowanej,
- populacji dziennej audytowanej,
- struktury demograficznej z zewnętrznego badania strukturalnego.

Sumując wagi tak zważonych aktywnych panelistów z danego dnia uzyskujemy populację używaną w danym Cyklu Produkcyjnym. Cykl Produkcyjny w przypadku Constant Panelu to jeden dzień.

Dzięki powyższemu możliwe jest osiągnięcie następujących korzyści:

- populacja aktualizowana co miesiąc,
- populacja wyznaczana w przód, na kolejne miesiące, zamiast używania populacji z zakończonego już okresu,
- wierne oddawanie trendów sezonowych,
- łagodniejsze przejścia pomiędzy falami badania strukturalnego.

4.5. Odwzorowanie liczby odsłon, czasu i wizyt dla witryn i aplikacji audytowanych (Algorytm FPA 2.0)

Jest to algorytm edytowania aktywności panelistów w taki sposób, aby w najwyższym możliwym stopniu aktywność ta odpowiadała wynikom pozyskanym z pomiaru site-centric (odsłony, wizyty, czas). Cel realizowany jest poprzez dodawanie lub odejmowanie aktywności panelistów na poszczególnych węzłach.

Właściwości algorytmu FPA:

- Odsłony, czas i wizyty dla witryn i aplikacji audytowanych są bliskie danym site-centric - różnice nie powinny przekraczać 1%.
- Suma arytmetyczna odsłon i czasu z serwisów danej domeny może być minimalnie mniejsza od wyniku podanego dla całej domeny. Różnica wynikać może z tego, że na serwisy o małej liczbie użytkowników w niektórych dniach, na niektórych platformach (pc-home, pc-work, phones, tablets), mogą nie zostać odwiedzone przez żadnego panelistę. W przypadku gdy liczba odsłon na danej platformie w danym dniu nie jest większa niż waga jednego constant panelisty (162), to dane o liczbie odsłon, wizytach i czasie nie są prezentowane dla tego serwisu na danej platformie (jest widoczne zero). Są one jednak uwzględniane w wynikach dla całej domeny, ponieważ na domenie liczba panelistów jest już niezerowa. W większości przypadków różnica ta nie przekracza 0,3% odsłon.
- Wizyty są liczone jako ciąg odsłon na danej domenie, więc suma wizyt z jej serwisów zawsze będzie istotnie wyższa niż liczba wizyt prezentowana na całej domenie. Wizyta na serwisie jest wizytą na danej domenie przechodzącą m.in. przez ten serwis. W czasie jednej wizyty na danej domenie użytkownik może odwiedzić wiele serwisów w tej domenie.
- Dane o odsłonach, czasie i wizytach na węzłach grup właścicielskich są sumą arytmetyczną tych wskaźników z węzłów podrzędnych (domen i aplikacji). Mogą się jednak różnić w bardzo małym stopniu ze względu na zaokrąglenia (sumowane są liczby niecałkowite, a w interfejsie są widoczne one w zaokrągleniu do liczb całkowitych). Różnice te nie będą jednak przekraczać liczby węzłów sumowanych razy liczba dni razy liczba platform (pc-home, pc-work, phones, tablets). Nawet dla grup o wielu węzłach różnice te nie przekraczają zwykle 10 odsłon.

4.6. Sposób estymacji liczby przeglądarek (EC Global)

Estimated Cookie to szacowana liczba cookies odwiedzających dany węzeł, w sytuacji gdyby nie istniało zjawisko ich kasowalności i wszystkie przeglądarki akceptowałyby cookies Gemiusa. Innymi słowy, jest to szacowana liczba przeglądarek.

EC Global uniemożliwia wpływanie na współczynnik AH (all hits)/GH (good hits) na poziomie witryny/grupy i jest odporny na modyfikacje wpływające na wielkość ruchu.

4.7. Metodologia produkcji wyników dla witryn niepodlegających audytowi site-centric

Software i Hardware panel

Wyniki witryn nieaudytowanych są estymacją opartą na panelu badawczym. Paneliści są rekrutowani poprzez ankietę wyświetlaną na wszystkich audytowanych w badaniu stronach. Każdy, kto po wypełnieniu kwestionariusza z danymi demograficznymi zainstaluje na swoim urządzeniu rozszerzenia NetPanel, zostaje częścią softwarepanelu. Panel telefoniczny jest także tworzony przez osoby, które otrzymały od nas telefon z zainstalowanym na stałe miernikiem badawczym (tzw. hardware panel). Dzięki panelowi software i hardware informacja o każdej wykonanej przez panelistę odsłonie jest do nas przesyłana wraz z informacją o jej adresie URL. Zbieramy również informację o czasie przez jaki otwarte są poszczególne strony internetowe w aktywnej karcie w przeglądarce.

Proces ważenia

Zebrane dane panelistów są poddawane procesowi ważenia. Wagi są nadawane w taki sposób, aby ich suma była zgodna z danymi o liczbie internautów z badania NetTrack firmy Kantar oraz aby struktura demograficzna przeważonych danych była zgodna ze strukturą z tego badania. W ważeniu demograficznym uwzględniamy cechy takie jak płeć, wiek, wykształcenie, wielkość miejscowości i województwo. Dodatkowo w procesie ważenia brane są pod uwagę warunki behawioralne, czyli aktywność panelistów. Staramy się tak ukształtować wagi, aby suma wag panelistów odwiedzających witryny poddane audytowi site-centric zbliżona była do liczby użytkowników wyliczonej na podstawie twardych danych zebranych w systemie site-centric. Dla wyliczania danych dla witryn niepodlegających audytowi site-centric używamy w ten sposób skorygowanych wag, aby zapewnić jak największą porównywalność liczby użytkowników pomiędzy witrynami audytowanymi i nieaudytowanymi.

Balance amendement

Należy wziąć pod uwagę to, iż w badaniu jest nadreprezentacja panelistów korzystających bardzo dużo z internetu. Z tego powodu obniżamy wagi panelistów bardzo aktywnych w internecie a podwyższamy wagi panelistów mało aktywnych. Proces ten nosi nazwę „balance amendement” i jest oparty na porównaniu liczby odsłon oszacowanej na podstawie przeważonego we wcześniejszych etapach ważenia panelu z liczbą odsłon pochodzącą z twardych danych site-centric na witrynach audytowanych. Dzięki tej części

naszej metodologii staramy się zapewnić jak największą porównywalność wskaźników aktywności (odstłon, czasu i wizyt) pomiędzy witrynami audytowanymi i nieaudytowanymi.

4. Dane w interfejsie mediapanel.gemius.com

Dane zaprezentowane w interfejsie obejmują okres od 1 września 2019 roku.

Dane od 1.09.2019 do 30.09.2020 to dane nieoficjalne i testowe, które powinny służyć jedynie w celu analiz podglądowych. W tym okresie danymi oficjalnymi były dane udostępniane w aplikacji gemiusExplorer i jedynie one powinny służyć jako oficjalne źródło informacji na temat zachowań użytkowników w Internecie.

Od dnia 1.10.2020 dane udostępniane w interfejsie online są danymi oficjalnymi.

5. Publikacja danych

Publikacja danych w interfejsie mediapanel.gemius.com jest procesem, który odbywa się w cyklach dziennych. Oznacza to, że każdego dnia roboczego dane są publikowane za poprzedni dzień. Jako że dane obliczane są na podstawie metodologii Constant Panel, publikacji podlegają jedynie okresy jednodniowe, a okresy dłuższe mogą być przez użytkownika dowolnie komponowane poprzez sumowanie okresów dziennych.

W standardowe dni robocze publikacja danych odbywa się do godziny 15:00. Oznacza to, że w interfejsie dane za dzień N powinny się pojawić do godz. 15:00 w dniu N+1.

W dni wolne od pracy (soboty, niedziele oraz dni ustawowo wolne) publikacja danych nie odbywa się. Dane za dzień/dni sprzed dnia wolnego od pracy są publikowane w następny dzień roboczy. W specyficznych okresach (np. okres Świąt Bożego Narodzenia) lub gdy w niektórych latach następuje kumulacja kilku dni wolnych od pracy (np. czwartek-niedziela), kalendarz publikacji będzie każdorazowo ustalany przez Gemius.

Publikacja danych dziennych poprzedzona jest weryfikacją po stronie Gemius. W przypadku stwierdzenia nieprawidłowości dopuszczalne są opóźnienia w publikacji, o czym Gemius będzie informował użytkowników interfejsu.