

Zastosowania systemów analityki internetowej do analizy efektywności kanałów ruchu i handlu elektronicznego

Artur Strzelecki

Wstęp

Analityka internetowa każdego roku wzbogaca się o nowe rozwiązania i metody pozyskiwania danych na temat zachowań użytkowników w Internecie. W niniejszym rozdziale zostaną przedstawione metody zbierania danych analitycznych i wyciągania z nich wniosków w dziedzinach, które nie zdobyły jeszcze dużej popularności i nie są powszechnie wykorzystywane. Na początku zostanie przedstawione w jaki sposób systemy analityki internetowej zbierają dane o użytkownikach oraz czym są kanały ruchu i jak należy je definiować. Następnie zostanie zaprezentowana analityka wielokanałowa oraz śledzenie transakcji w sklepach internetowych, tradycyjne i ulepszone. Na koniec przedstawione zostaną aktualne problemy analityki internetowej w zakresie fałszowania danych dostępnych w systemach analitycznych.

Użytkownik w systemie analityki internetowej

Systemy analityki internetowej zbierają informacje o odwiedzających użytkowników z różnych źródeł. Wyróżniamy trzy główne sposoby zbierania danych z różnych źródeł¹. Do tych źródeł należą dane zbierane przez serwer internetowy, dane zbierane na podstawie sesji oraz dane zbierane na podstawie interpretacji plików JavaScript. Każdy ze sposobów ma swoje dodatnie i ujemne cechy, którymi się charakteryzuje. Każdą z metod można scharakteryzować następująco.

Pierwszym typem zbieranych danych, są informacje pochodzące z oprogramowania zainstalowanego na serwerze do wyświetlania stron internetowych. Systemy analityki internetowej takie jak Awstat, Webalizer opierają się na danych pochodzących z serwera internetowego. Analizują logi z informacjami o hitach na serwerze. Hit zawiera informacje o adresie IP, dacie, godzinie, odwiedzanym adresie, kodzie odpowiedzi serwera oraz nazwie User Agent, która reprezentuje przeglądarkę internetową.

Przykładowy hit z danymi dla osoby korzystającej z przeglądarki internetowej będzie wyglądał następująco:

```
109.239.235.221 - - [10/Sep/2015:02:10:18 +0200] "GET / HTTP/1.1" 200 11991 "-"  
"Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_2) AppleWebKit/537.36 (KHTML, like  
Gecko) Chrome/34.0.1847.131 Safari/537.36"
```

¹ K. Nakatani, T. Chuang, (2011), "A web analytics tool selection method: an analytical hierarchy process approach", Internet Research, Vol. 21 Iss 2 ss. 171 – 186

Przykładowy hit z danymi dla programu automatycznie odwiedzającego strony internetowe internetowej będzie wyglądał następująco:

```
68.180.228.238 - - [10/Sep/2015:01:55:54 +0200] "GET / HTTP/1.1" 304 275 "-"  
"Mozilla/5.0 (compatible; Yahoo! Slurp; http://help.yahoo.com/help/us/ysearch/slurp)"
```

Pełny zbiór wszystkich hitów zawiera kompletne informacje o odwiedzinach na stronie internetowej. Niestety w tym zbiorze znajdują się zarówno odwiedziny użytkowników, czyli osób korzystających z urządzenia za pomocą, którego przeglądają strony internetowe, jak również wszystkich programów, które w automatyczny sposób odwiedzają i pobierają zawartość strony internetowej. Stąd nie sposób odróżnić, kto jest rzeczywistą osobą korzystającą ze strony internetowej, a automatycznym programem odczytującym zawartość tej samej strony, ponieważ wartość ostatniego pola User Agent można dowolnie modyfikować.

Drugim sposobem zbierania danych jest zapisywanie danych o użytkownikach, którzy korzystają ze stron internetowych tworzących dla każdego z nich własne sesje. Analityka internetowa oparta o sesje, dotyczy stron, które są dostępne po zalogowaniu. Na przykład systemy bankowości elektronicznej, badają ruch użytkowników w swoim systemie bankowości. Po zaprzestaniu korzystania z systemu i wylogowaniu się z niego sesja jest zamykana.

Trzecia metoda zbierania danych oparta jest o informacje wysyłane przez przeglądarki internetowe do centralnego serwera z danymi. Systemy analityki internetowej oparte o znaczniki JavaScript są obecnie najbardziej popularnymi. Do każdej strony internetowej dodawany jest fragment kodu JavaScript, który przy każdym odwiedzeniu tej strony przez przeglądarkę internetową obsługującą ten język przekazuje zebrane dane do centralnego serwera z danymi analitycznymi.

Przykład fragmentu kodu, który dodany do strony internetowej pozwoli zbierać dane o odwiedzających użytkowników, dla usługi Google Analytics zainstalowanej na stronie Uniwersytetu Ekonomicznego w Katowicach wygląda następująco:

```
<script>  
(function(i,s,o,g,r,a,m){i['GoogleAnalyticsObject']=r;i[r]=i[r]||function(){  
(i[r].q=i[r].q||[]).push(arguments)},i[r].l=1*new Date();a=s.createElement(o),  
m=s.getElementsByTagName(o)[0];a.async=1;a.src=g;m.parentNode.insertBefore(a,m)  
})(window,document,'script','/www.google-analytics.com/analytics.js','ga');  
ga('create', 'UA-9386373-1', 'auto');  
ga('send', 'pageview');  
</script>
```

Po otwarciu strony internetowej Uniwersytetu Ekonomicznego w Katowicach zawierającej powyższy kod, w pamięci przeglądarki internetowej zostanie utworzony plik cookie, natomiast informacje na temat przeglądarki internetowej zostaną przekazane do centralnego systemu zarządzającego danymi.

Zawartość pliku cookie składa się z ciągu liczb, z których każda zawiera informacje o numerze kolejnej wizyty na stronie oraz przeglądarce internetowej. Przykładowa zawartość tego pliku wygląda następująco
GA1.3.1917153934.1441876334

Zestaw informacji wysłanych centralnego serwera gromadzącego dane na przykładzie Google Analytics wygląda następująco:

Sent beacon:

```
v=1&_v=j38d&a=1705543444&t=pageview&_s=1&dl=http%3A%2F%2Fwww.ue.katowice.pl%2Fno_cache.html&ul=pl&de=UTF-8&dt=Uniwersytet%20Ekonomiczny...&_u=AACAAEABI~&jid=&cid=1917153934.1441876334&tid=UA-9386373-1&z=109053461
```

```
_jl          (&jid)
adSenseId   (&a) 1705543444
apiVersion  (&v) 1
clientId    (&cid) 1917153934.1441876334
encoding     (&de) UTF-8
flashVersion (&fl) 18.0 r0
hitType      (&t) pageview
javaEnabled  (&je) 1
language     (&ul) pl
location     (&dl) http://www.ue.katowice.pl/no_cache.html
screenColors (&sd) 24-bit
screenResolution (&sr) 1440x900
title        (&dt) Uniwersytet Ekonomiczny w Katowicach: Strona główna
trackingId   (&tid) UA-9386373-1
viewportSize (&vp) 1440x801
```

Powyższe dane zostały uzyskane za pomocą rozszerzenia Google Analytics Debugger dostępnego w przeglądarce Google Chrome. Wynika z tego, że użytkownikiem w systemach analityki internetowej jest przeglądarka internetowa z zainstalowanym plikiem cookie.

Bez względu na to z jakiego sposobu zbierania danych się korzysta oraz jaki wykorzystuje się system do analityki internetowej, każdy taki system na podstawie zebranych danych powinien pomóc uzyskać informacje i pomóc wyciągnąć wnioski, na podstawie których mogą być podejmowane dalsze decyzje.

Źródła ruchu

Każdy użytkownik odwiedzający stronę internetową, która śledzi ruch użytkowników odwiedza ją przychodząc z określonego źródła^{2 3}. Źródła ruchu dzielą się na sześć głównych kanałów z których przychodzą użytkownicy odwiedzający stronę internetową. Do tych kanałów ruchu należą:

1. Wyszukiwarki internetowe
2. Sieci społecznościowe
3. Odesłania z innych stron
4. Wejścia bezpośrednie
5. Reklama graficzna

² B. Plaza, (2009), "Monitoring web traffic source effectiveness with Google Analytics", Aslib Proceedings, Vol. 61 Iss 5 ss. 474 - 482

³ H. Pakkala, K. Presser, T. Christensen, (2012) Using Google Analytics to measure visitor statistics: The case of food composition websites. International Journal of Information Management 32, 504-512.

6. E-mail

Z reguły największy wolumen ruchu internetowego dostarczają wyszukiwarki internetowe. Użytkownicy korzystają z nich aby odnaleźć docelowe strony internetowe. Usługi badające ruch internetowy mają utworzoną listę hostów, które uznane są za źródła ruchu zaliczane jak wyszukiwarki. Takie źródło oprócz nazwy hosta ma zdefiniowany parametr, którego wartość jest słowem kluczowym wprowadzonym do wyszukiwarki. Po wprowadzeniu zapytania do wyszukiwarki, zwrócona lista rezultatów zawiera hiperłącza do stron docelowych, jednak przejście przez takie hiperłącze nie będzie traktowane jak przejście z innej strony internetowej, tylko skorzystanie z wyszukiwarki internetowej. Np. w usłudze Google Analytics poniższe strony internetowe, są zdefiniowane jako wyszukiwarki wraz z parametrem, po którym następuje wyszukanie:

Tabela nr 1

| Wyszukiwarka | Adres wyszukiwarki | Parametr |
|---------------------|---|-----------------|
| Daum | http://www.daum.net/ | q |
| Eniro | http://www.eniro.se/ | search_word |
| Naver | http://www.naver.com/ | query |
| Google | http://www.google.com | q |
| Yahoo | http://www.yahoo.com/ | p |
| MSN | http://www.msn.com/ | q |
| Bing | http://www.bing.com/ | q |
| AOL | http://www.aol.com/ | query |
| AOL | http://www.aol.com/ | encquery |
| Lycos | http://www.lycos.com/ | query |
| Ask | http://www.ask.com/ | q |
| Altavista | http://www.altavista.com/ | q |
| Netscape | http://search.netscape.com/ | query |
| CNN | http://www.cnn.com/SEARCH/ | query |
| About | http://www.about.com/ | terms |
| Mamma | http://www.mamma.com/ | query |
| Alltheweb | http://www.alltheweb.com/ | q |
| Voila | http://www.voila.fr/ | rdata |
| Virgilio | http://search.virgilio.it/ | qs |
| Live | http://www.bing.com/ | q |
| Baidu | http://www.baidu.com/ | wd |
| Alice | http://www.alice.com/ | qs |
| Yandex | http://www.yandex.com/ | text |
| Najdi | http://www.najdi.org.mk/ | q |
| AOL | http://www.aol.com/ | q |
| Mama | http://www.mamma.com/ | query |
| Seznam | http://www.seznam.cz/ | q |
| Search | http://www.search.com/ | q |

| | | |
|-------------------|---|--------|
| Wirtualana Polska | http://www.wp.pl/ | szukaj |
| O*NET | http://online.onetcenter.org/ | qt |
| Szukacz | http://www.szukacz.pl/ | q |
| Yam | http://www.yam.com/ | k |
| PCHome | http://www.pchome.com/ | q |
| Kvasir | http://www.kvasir.no/ | q |
| Sesam | http://sesam.no/ | q |
| Ozu | http://www.ozu.es/ | q |
| Terra | http://www.terra.com/ | query |
| Mynet | http://www.mynet.com/ | q |
| Ekolay | http://www.ekolay.net/ | q |
| Rambler | http://www.rambler.ru/ | words |

Źródło: Dostosowanie źródeł w bezpłatnych wynikach wyszukiwania,
<https://support.google.com/analytics/answer/2795821?hl=pl>

Wybór odnośnika znajdującego się na liście wyników zostanie zapisany jako ruch z wyszukiwarki organicznej. Organicznej, czyli wyniki są przygotowane na podstawie algorytmu wewnętrznego wyszukiwarki, w żaden sposób nie związanego z linkami sponsorowanymi.

Wybór odnośnika oznaczonego jako link sponsorowany lub reklamowy zostanie zapisany jak ruch z wyszukiwarki płatnej. Źródło ruchu wyszukiwarka jest podzielone w ten sposób na dwie podgrupy: organiczną i sponsorowaną (płatną).

Ruch z sieci społecznościowych to przejścia do strony internetowej z jednej z wielu znanych sieci społecznościowych. Podobnie jak w przypadku wyszukiwarek internetowych, systemy analityki internetowej posiadają wbudowane listy adresów hostów, które są traktowane jako źródło ruchu pochodzące z sieci społecznościowej. Użytkownicy sieci społecznościowych spędzają w nich relatywnie dużo czasu i korzystają z wielu hiperłączy, które są tam umieszczane przez innych użytkowników. Przejście przez takie hiperłącze jest zaliczane do ruchu z sieci społecznościowych.

Przykładowa lista portali społecznościowych, która jest zdefiniowana jako ruch z sieci społecznościowych w usłudze Google Analytics znajduje się poniżej:

- Facebook
- Twitter
- Blogger
- GoldenLine
- LinkeInd
- Pocket
- Yelp
- nk.pl
- LiveJournal

Poniższa lista to sieci społecznościowe, które są partnerami usługi Google Analytics i dostarczają do niej pełny zakres danych:

- AllVoices

- Google Groups
- Reddit
- Delicious
- Google+
- SodaHead
- Diigo
- Hatena
- VKontakte
- Disqus
- Meetup
- Yaplog

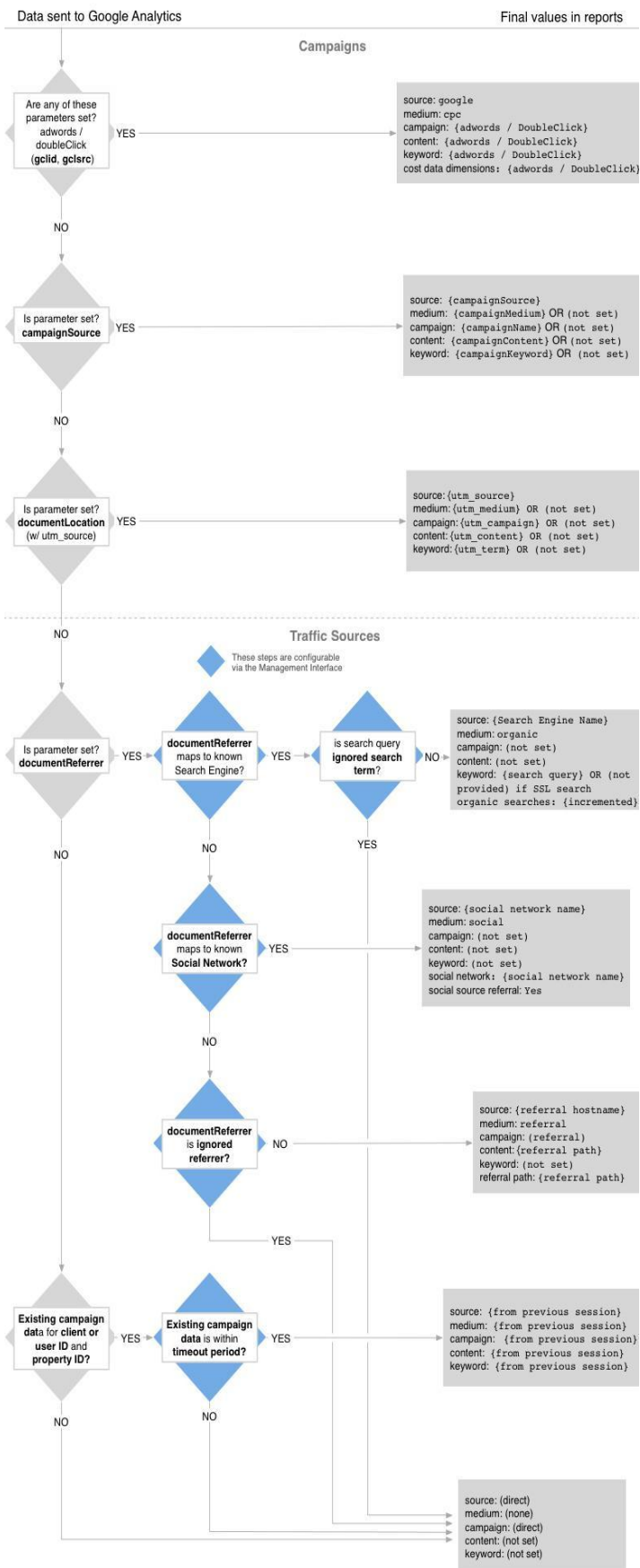
Ruch z odesłania to ruch z innych domen internetowych niż badana strona internetowa, a jednocześnie ten ruch nie pochodzi ze znanej wyszukiwarki internetowej lub ze znanej sieci społecznościowej. Wartością określającą dokładne źródło ruchu jest nazwa domeny internetowej.

Wejścia bezpośrednie to ruch na stronie internetowej, który rozpoczął się bezpośrednio, nie poprzedziła go wizyta na innej stronie internetowej. Wejścia bezpośrednie powstają, gdy użytkownik bezpośrednio wpisze adres strony internetowej w pasku adresu przeglądarki internetowej, skorzysta z zakładek obecnych w przeglądarce internetowej, skorzysta z historii odwiedzanych stron zapisanych w przeglądarce lub gdy skorzysta z zewnętrznego programu w którym zostało uruchomione hiperłącze i wywołano okno przeglądarki internetowej. Takimi zewnętrznymi programami mogą być klienci poczty elektronicznej jak Microsoft Outlook lub Thunderbird, edytory tekstu lub arkusze kalkulacyjne oraz programy wyświetlające zawartość dokumentów pdf, np. Adobe Acrobat.

Kanał ruchu określany jako reklama graficzna zawiera informacje o ruchu prowadzącym do strony internetowej poprzez kliknięcie baniera reklamowego w graficznej postaci. Graficzna reklama znajduje się zazwyczaj w treści strony internetowej, ale nie jest traktowana jako ruch z odesłania, ponieważ fizycznie nie należy on tej strony, a jedynie jest wyświetlana z dedykowanego serwera reklamowego. Ruch z reklama graficznych jest rozpoznawany na podstawie dodatkowych informacji o źródle ruchu dostarczanych przez serwery reklamowe. Reklama graficzna może też być powiązana z remarketingiem, czyli z docieraniem przekazu reklamowego na podstawie już wcześniej zainstalowanych plików cookie.

Ruch internetowy z wiadomości email może zostać zaliczony do tego kanału, gdy wiadomości email są wysyłane przez dedykowane systemy do email marketingu. Listy wysyłane poprzez taki system posiadają wbudowane informacje w nagłówek listu, na podstawie których system analityczny potrafi rozpoznać, że ten rodzaj ruchu pochodzi z odczytania wiadomości w programie pocztowym. Zapisanie ruchu do tego kanału jest nadrzędne i nie ma znaczenia czy odczytanie wiadomości oraz kliknięcie hiperłącza nastąpiło w dedykowanym programie pocztowym jak Microsoft Outlook i Thunderbird czy w kliencie poczty online jak Gmail lub Yahoo.

Na rysunku 1 przedstawiono schemat blokowy ilustruje proces rozpoznawania źródła ruchu w usłudze Google Analytics.



Rysunek nr 1, Kampanie i źródła wizyt,

Źródło: <https://support.google.com/analytics/answer/6205762?hl=pl>

Użytkownicy wykorzystują wyżej wymienione źródła ruchu, aby dotrzeć do strony internetowej. Odwiedzając stronę internetową mogą wykonać cele, które są mierzone w usłudze analitycznej jako konwersje. Konwersja to wykonanie założonego celu przez użytkownika odwiedzającego stronę internetową.

Przykładowymi celami do wykonania przez użytkownika mogą być:

w zakresie sprzedaży:

1. złożenie zamówienia rozumiane jako zamówienie wstępne lub dokonany zakup,
2. zakup towaru rozumiany jako dokonanie zakupu online,

w zakresie pozyskania:

1. utworzenie konta rozumiane jako rejestrację lub założenie konta,
2. przesłanie treści rozumiane jako przesłanie pliku,

w zakresie uzyskania informacji:

1. polecenie treści znajomemu jako wysłanie linku e-mailem,
2. przeczytanie recenzji poprzez wyświetlenie opinii i oceny,
3. wyświetlenie kalendarza lub otwarcie listy nadchodzących wydarzeń,
4. pobranie lub wydrukowanie broszury informacyjnej,

w zakresie zaangażowania:

1. odtworzenie multimediiów poprzez interaktywne odtwarzacze multimedialne np. filmów, pokazów slajdów lub prezentacji produktów,
2. udostępnienie w sieci społecznościowej lub połączenie z nią,
3. prośba o okres próbny jako oczekiwanie na wersję testową lub próbną.

Konwersja czyli wykonanie jednego z wyżej określonych celów odbywa się na stronie internetowej i przypisane jest do źródła ruchu z którego przybył użytkownik. Jednak wielu użytkowników, gdy mają wykonać cel na stronie, może ją odwiedzić więcej niż jeden raz i uczynić z więcej niż jednego źródła. Szczególnie różne źródła ruchu mogą pojawić się w sytuacji, gdy użytkownik chce zakupić jakiś produkt w sklepie internetowym.

Systemy analityki internetowej domyślnie przypisują konwersję do ostatniego źródła ruchu, z którego skorzystał użytkownik odwiedzający stronę internetową. Te domyślne ustawienia będą poprawne jeśli po wejściu na stronę internetową użytkownik korzystał tylko z jednego, tego samego źródła ruchu. W tym przykładzie konwersję można zakwalifikować do tego źródła ruchu. Co jednak zrobić w sytuacji, jeżeli historia odwiedzin strony internetowej jest bardziej złożona i do wykonania konwersji pod drodze były odwiedziny z kilku innych źródeł ruchu?

Przypisanie konwersji do ostatniego źródła ruchu wydają się być prawidłowe, to jednak nie ono musiało zapoczątkować odwiedzin na tej stronie. Narzędziem, które pozwala uzyskać szerszy obraz, które źródła ruchu mają udział przy konwersjach dokonywanych na witrynie internetowej są tzw. ścieżki wielokanałowe lub analityka wielokanałowa⁴

Analityka wielokanałowa

Analityka wielokanałowa stara się odpowiedzieć na pytanie jakie poprzednie źródła ruchu przyczyniły się do wykonania konwersji na stronie internetowej. Analityka wielokanałowa polega na analizowaniu ścieżek konwersji. Ścieżka konwersji to sekwencja interakcji, czyli wejść z tych samych bądź różnych kanałów ruchu i poprzedza ona wykonanie konwersji.

⁴A. Kaushik, (2010) Web Analytics 2.0. Świadome rozwijanie witryn internetowych, Helion, Gliwice.

Analityka wielokanałowa działa w oparciu o analizę wielu kanałów ruchu dla jednego użytkownika w określonym przedziale czasu. Przeważnie dotyczy ona okresu ostatnich 30 dni przed wykonaniem konwersji, natomiast systemy analityki internetowej pozwalają ten zakres modyfikować na przykład w przedziale od 1 do 90 dni.

Dane ze ścieżek konwersji obejmują informacje o ruchu z każdego kanału ruchu internetowego jak:

1. Wyszukiwarki internetowe
2. Sieci społecznościowe
3. Witryny odsyłające
4. Wejścia bezpośrednie
5. Reklama graficzna
6. Newslettery e-mail

Narzędzia badające analitykę wielokanałową potrafią zilustrować z jakich kanałów korzystali użytkownicy w okresie poprzedzających konwersję, czyli udzielić odpowiedzi na pytanie jak i kiedy różne kanały wpłynęły na daną konwersję. Ilustrują także ile czasu mija, zanim dojdzie do konwersji, jaka jest długość ścieżki poprzedzających konwersję i jakie są najważniejsze trasy, które pokonują użytkownicy.

Na ścieżce konwersji każdy kanał ruchu internetowego może być traktowany jako jeden z trzech możliwych wariantów:

1. Ostatnia interakcja to kanał, z którego przyszedł użytkownik a następnie wykonał konwersję.
2. Interakcja pośrednia, to dowolny kanał, który znalazł się na ścieżce konwersji ale nie jest ani ostatni ani pierwszy na niej.
3. Pierwsza interakcja to pierwszy kanał ruchu internetowego, który znalazł się na ścieżce konwersji.

Narzędzie badające analitykę wielokanałową powinno potrafić ustalić jaki wymiar konwersji został uzyskany dzięki interakcjom wspomaganym oraz pokazać wszystkie kanały ruchu, które brały udział przy tych konwersjach. Podobnie powinno być z pierwszą i ostatnią interakcją. Narzędzie analityki internetowej prawidłowo rozpozna całą ścieżkę poprzedzającą konwersję, będzie posiadało informacje, który kanał był pierwszy, który był ostatni na tej ścieżce i jakie kanały brały udział w jej tworzeniu.

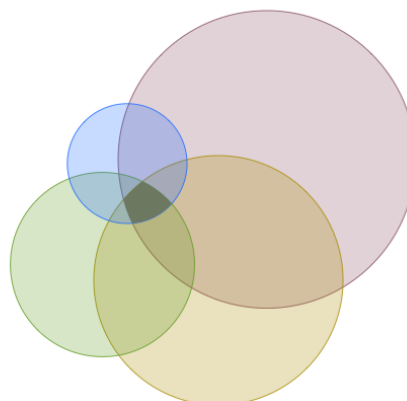
Rysunek nr obrazuje, że przy dokonywaniu konwersji na stronie internetowej mają miejsce interakcje pochodzące z różnych kanałów. Suma wartości procentowych przy każdym kanale, przewyższa 100% co oznacza, że wiele kanałów ruchu internetowego, znajduje się na drodze do wykonania konwersji.

Zobrazowanie konwersji wielokanałowej

Odszetek ścieżek konwersji, które zawierają kombinacje poniższych kanałów. Wybierz maksymalnie cztery kanały.

| Kanał | % łącznej liczby konwersji |
|---|----------------------------|
| <input checked="" type="checkbox"/> Odesłania | 63,14% |
| <input checked="" type="checkbox"/> Wejścia bezpośrednie | 44,37% |
| <input checked="" type="checkbox"/> Bezpłatne wyniki wyszukiwania | 24,23% |
| <input checked="" type="checkbox"/> Płatne wyniki wyszukiwania | 10,24% |
| <input type="checkbox"/> Sieć społecznościowa | 1,71% |

Odesłania & Wejścia bezpośrednie & Bezpłatne wyniki wyszukiwania & Płatne wyniki wyszukiwania: 0.68% (2)



Rysunek nr 2, Zobrazowanie konwersji wielokanałowej

Źródło: Google Analytics, opracowanie własne

Narzędzie analityki internetowej powinno także zilustrować najważniejsze trasy, które pokonują użytkownicy na drodze do konwersji. Będą to sekwencje między kanałami. Te dane powinny także pomóc odkryć powtarzające się wzorce, które podpowiadają, które kanały ruchu internetowego są najefektywniejsze.

Rysunek nr 3 ilustruje przykład sekwencji kanałów, które pojawiają się na drodze do wykonania konwersji na stronie internetowej.

| Ścieżka grupowania kanałów ścieżek wielokanałowych ? | Konwersje |
|--|-------------|
| 1. Odesłania → Wejścia bezpośrednie | 21 (15,56%) |
| 2. Odesłania × 2 | 10 (7,41%) |
| 3. Odesłania → Wejścia bezpośrednie × 2 | 9 (6,67%) |
| 4. Bezpłatne wyniki wyszukiwania → Wejścia bezpośrednie | 9 (6,67%) |
| 5. Odesłania → Wejścia bezpośrednie → Odesłania | 6 (4,44%) |
| 6. Wejścia bezpośrednie × 2 | 4 (2,96%) |
| 7. Bezpłatne wyniki wyszukiwania → Wejścia bezpośrednie × 2 | 4 (2,96%) |
| 8. Odesłania × 2 → Wejścia bezpośrednie → Odesłania | 3 (2,22%) |
| 9. Odesłania → Płatne wyniki wyszukiwania → Odesłania → Wejścia bezpośrednie | 2 (1,48%) |
| 10. Odesłania × 2 → Wejścia bezpośrednie | 2 (1,48%) |

Rysunek nr 3, Ścieżki wielokanałowe

Źródło: Google Analytics, opracowanie własne

Analityka wielokanałowa ilustruje także, jak długie były sekwencje kanałów, które prowadziły do wykonania konwersji. Ścieżki mogą posiadać długość od jednego do kilkunastu kanałów znajdujących się na drodze do wykonania konwersji. Podobnie jeśli chodzi o analizę upływu czasu przed dokonaniem konwersji. Narzędzie do analityki powinno pokazać jak długo trwało przechodzenie pomiędzy kanałami.

Po ustaleniu jak przebiegają ścieżki użytkowników i które sekwencje kanałów ruchu biorą w nich udział należy ustalić, które z tych kanałów będą miały udział w konwersji i w jakiej części. Służą do tego modele atrybucji.

Modele atrybucji

Model atrybucji jest regułą lub zbiorem reguł, które określają sposób przypisywania konwersji do określonych kanałów ruchu znajdujących się na ścieżce wielokanałowej

Do klasycznych, domyślnych modeli atrybucji należą:

1. Ostatnia interakcja.
2. Ostatnie kliknięcie niebezpośrednie.
3. Ostatnie kliknięcie w link sponsorowany.
4. Pierwsza interakcja.
5. Liniowy.
6. Rosnący.
7. Uwzględnienie pozycji.

Model atrybucji ostatniej interakcji przydzieli 100% udziału w konwersji ostatniemu kanałowi, z którego użytkownik wszedł na stronę zanim dokonał konwersji. Model ostatniej interakcji można stosować, gdy działania nakłaniające do konwersji mają miejsce w momencie zakupu lub gdy konwersja bazuje na działaniach pomijających fazę namysłu.

Model atrybucji ostatnie kliknięcie niebezpośrednie ignoruje ruch z kanału ruchu bezpośredniego i przydziela 100% udziału w konwersji ostatniemu kanałowi, z którego użytkownik wszedł na stronę zanim dokonał konwersji. To domyślny model stosowany w systemach analityki internetowej.

Model atrybucji ostatnie kliknięcie w link sponsorowany przydzieli 100% udziału w konwersji ostatniej reklamie sponsorowanej, którą kliknął użytkownik przed dokonaniem konwersji. Model ten można stosować aby zidentyfikować i przypisać udział linkom sponsorowanym, które doprowadziły do największej liczby konwersji.

Model atrybucji pierwszej interakcji przydzieli 100% udziału w konwersji pierwszemu kanałowi, z którym użytkownik wszedł w interakcję. Model ten można stosować na stronach internetowych, gdzie buduje się początkową świadomość.

Model atrybucji liniowy przypisze taki sam udział wszystkim kanałom interakcji na ścieżce wielokanałowej prowadzącej do konwersji. Model ten można stosować kiedy użytkownik jest tak samo ważny, bez względu na to z którego kanału ruchu przychodzi.

Model atrybucji rosnący jest oparty o rozkład czasowy. Ten model opiera się na koncepcji rozkładu wykładniczego i przypisuje największy udział kanałom ruchu znajdującym się najbliżej konwersji. Model można stosować, wtedy gdy dochodzi do interakcji i konwersji w dniach, w których prowadzone są działania nakłaniające do wykonania konwersji. Mniejsze znaczenie będą miały wtedy interakcje, który odbyły się wcześniej, ale nie doprowadziły do konwersji.

Model atrybucji uwzględniający pozycję łączy ze sobą cechy modeli pierwszej interakcji oraz ostatniej interakcji. Nie przypisuje całego udziału pierwszej lub ostatniej interakcji, natomiast jest on dzielony pomiędzy te interakcje. Często propozycją jest przypisanie pierwszej i

ostatniej interakcji po 40% udziału w konwersji, natomiast pozostałe 20% przypisywane jest kanałom ruchu, które są pośrednie. Model można używać wtedy, gdy najważniejszymi działaniami prowadzącymi do konwersji są pierwszy kontakt oraz etap prowadzący do konwersji.

Narzędzie analityki internetowej powinno umożliwiać również tworzenie własnych modeli atrybucji, choć te, które są tzw. klasycznymi modelami atrybucji są najczęściej wystarczające. W zależności od zastosowanego modelu atrybucji liczba konwersji przypisanych do każdego źródła ruchu internetowego będzie się różniła.

Kanał ruchu, który w największym stopniu inicjuje ścieżki konwersji, będzie miał wyższą wartość w modelu atrybucji pierwsza interakcja niż w modelu ostatnia interakcja.

Rysunek nr 4 przedstawia porównanie trzech modeli atrybucji (pierwsza interakcja, ostatnie kliknięcie niebezpośrednie, ostatnia interakcja) w tym samym okresie czasu.

| Grupowanie kanałów ścieżek wielokanałowych | Wydatki (dla wybranego okresu) | Konwersje i CPA | | | | | |
|--|-----------------------------------|---------------------|----------|-------------------------------------|----------|---------------------|----------|
| | | Pierwsza interakcja | | Ostatnie kliknięcie niebezpośrednie | | Ostatnia interakcja | |
| | | Konwersje | CPA | Konwersje | CPA | Konwersje | CPA |
| 1. Wejścia bezpośrednie | — | 19 705,00 (39,95%) | — | 10 876,00 (22,05%) | — | 26 540,00 (53,80%) | — |
| 2. Bezplatne wyniki wyszukiwania | — | 14 552,00 (29,50%) | — | 16 807,00 (34,07%) | — | 9 689,00 (19,64%) | — |
| 3. Odesłania | — | 10 104,00 (20,48%) | — | 13 873,00 (28,12%) | — | 9 482,00 (19,22%) | — |
| 4. (inne) | — | 2 299,00 (4,66%) | — | 3 730,00 (7,56%) | — | 1 291,00 (2,62%) | — |
| 5. Płatne wyniki wyszukiwania | 13 828,66 zł | 1 441,00 (2,92%) | 9,60 zł | 1 648,00 (3,34%) | 8,39 zł | 1 038,00 (2,10%) | 13,32 zł |
| 6. Sieć społecznościowa | — | 1 138,00 (2,31%) | — | 2 118,00 (4,29%) | — | 1 218,00 (2,47%) | — |
| 7. Reklama displayowa | 3 352,30 zł | 90,00 (0,18%) | 37,25 zł | 277,00 (0,56%) | 12,10 zł | 71,00 (0,14%) | 47,22 zł |

Rysunek nr 4, Porównanie modeli atrybucji

Źródło: Google Analytics, opracowanie własne

Każdy z modeli atrybucji przypisał inne udziały w liczbie dokonanych konwersji poszczególnym źródłom ruchu. W powyższym przykładzie stosując model atrybucji pierwszej interakcji, największy udział w konwersji ma kanał ruchu bezpośredniego. W modelu ostatnie kliknięcie niebezpośrednie, największy udział ma ruch z wyszukiwarki internetowej. Trzeci z wyświetlonych modeli, ostatnia interakcja ponownie największy udział w konwersjach przypisuje do kanału ruchu bezpośredniego.

Modele atrybucji można zestawiać z kosztami prowadzenia działań w różnych kanałach ruchu internetowego.

Systemy analityki internetowej pozwalają zbierać informacje o kosztach z różnych źródeł. Usługa Google Analytics jeśli jest również połączona z usługą linków sponsorowanych Google AdWords automatycznie gromadzi informacje o kosztach kanału linków sponsorowanych. Pozostałe informacje o kosztach można zaimportować z wykorzystaniem importu danych i plików tekstowych np. plików rozdzielanych przecinkami jak CSV.

Na rysunku nr 4 dwa kanały ruchu mają również przypisaną informację o koszcie jaki został poniesiony, aby te kanały mogły przynieść konwersję. Wartość kosztu podzielona przez liczbę konwersji pozwoli ustalić średni koszt uzyskania konwersji. Koszt pozyskania konwersji w każdym modelu atrybucji będzie inny ponieważ stały koszt będzie podzielony przez zmienną liczbę konwersji.

Pliki z danymi o kosztach powinny być zawsze zgodne ze specyfikacją narzędzia analityki internetowej. Z reguły taki zestaw danych powinien zawierać informację o źródle kanału ruchu, liczbę rozpoczętych interakcji oraz wartość kosztu. Dane powinny być segmentowane według dat, wtedy wszystkie miary będą jednostkowe w skali jednego dnia i przypisane do określonej daty.

Przykład zestawu danych do importu kosztów został przedstawiony w tabeli nr 2:

Tabela nr 2. Zestaw danych do importu kosztów

| Data | Medium | Źródło | Koszt (w PLN) | Kliknięcia |
|----------|-----------|----------|---------------|------------|
| 20150901 | odesłanie | ceneo.pl | 124,91 | 291 |
| 20150902 | odesłanie | ceneo.pl | 104.45 | 249 |
| 20150903 | odesłanie | ceneo.pl | 75.09 | 179 |

Źródło: Opracowanie własne.

Po dodaniu do systemu analityki internetowej informacji o dodatkowych kosztach poniesionych, które zostaną przypisane do właściwego kanału ruchu, system udostępni informacje o wskaźniku ROAS. ROAS to łączna kwota przychodów podzielona przez łączną kwotę wydatków. Na przykład ROAS w wysokości 500% oznacza, że każde 1PLN przychodu wymagał wydatka 0,20 PLN na ten kanał ruchu.

Tradycyjne śledzenie e-commerce

Wcześniej zostało przedstawione pojęcie celu, który może być śledzony w ramach systemu analityki internetowej. Kilka z przykładowych dotyczyło zachowań użytkowników w sklepach internetowych, jak złożenie zamówienia bądź dokonanie zakupu.

Właściwe śledzenie celów na stronie internetowej dostarczy informacji, jaka część użytkowników realizuje te cele w obrębie strony internetowej i jak wyglądają najczęstsze ścieżki dojścia do realizacji celu. Dotyczy to zarówno nawigacji wewnątrz strony internetowej i przechodzenia do kolejnych etapów realizacji celu, jak również ścieżki wielokanałowej, pokazującej, jaką często do tej pory użytkownik odwiedzał witrynę internetową i jakie źródła ruchu w tym uczestniczyły.

Do każdego zrealizowanego celu można przypisać wartość, dzięki temu można oszacować wielkość przychodu na podstawie zrealizowanych konwersji. Jednak taka wiedza, wyłącznie o przychodzie może nie być wystarczająca, aby prawidłowo analizować procesy zachodzące w sklepie internetowym.

Śledzenie celów w systemie analityki internetowej jeśli zostanie skonfigurowane w postaci dodatkowego modułu opartego o dane typowe dla transakcji w sklepie internetowym dostarczy lepszych jakościowo informacji do tego systemu. Powinno ono umożliwić zapisanie szeregu dodatkowych danych na temat wykonanej przez użytkownika konwersji, które byłyby zarejestrowane przez moduł śledzenia celów.

Dobrze rozwinięte systemy analityki internetowej pozwalają gromadzić i analizować dane o zakupach i transakcjach w sklepach internetowych⁵. Tego rodzaju dane umożliwiają analizowanie zakupów w witrynie internetowej lub w aplikacji. Znajdą się w nich informacje o produktach, transakcjach, średniej wartości zamówienia, współczynniku konwersji, czasie w jakim dokonywane są zakupy oraz inne dane.

System analityczny śledzący transakcje e-commerce pozwala sprawdzać, co kupują użytkownicy w sklepie internetowym lub aplikacji, m.in. taki system powinien udostępniać informacje o:

1. Produktach czyli jakie produkty są sprzedawane, w jakich ilościach oraz jak wpływa to na przychód.
2. Transakcjach czyli informacje dla poszczególnych transakcji jak przychody, podatek, koszty wysyłki oraz liczbie sprzedanych produktów lub usług.
3. Czasie do zakupu czyli liczbie dni oraz liczbie niezbędnych wizyt które poprzedziły realizację zakupu, licząc od czasu, kiedy użytkownik pierwszy raz odwiedził sklep internetowy aż do zakończonej transakcji.

Na podstawie takich informacji można dowiedzieć się:

1. Który asortyment sprzedaje się najlepiej oraz które produkty są najlepiej dopasowane do oczekiwań użytkowników i jakie kanały ruchu doprowadziły do najwyższego współczynnika konwersji.
2. Jaki jest wolumen przychodów oraz liczba produktów, które biorą udział w jednej transakcji. Jeśli liczba produktów jest zbyt niska, to na podstawie tych danych można rozważyć propozycję niższej ceny, przy zamówieniu większej ilości towarów lub rozważyć pokrycie kosztów przesyłki towaru po przekroczeniu przez użytkownika ustalonej wartości minimalnej transakcji.
3. Ile czasu oraz ile odwiedzin musi wykonać użytkownik nim podejmie ostateczną decyzję o zakupie i dokona konwersji. Jeśli cykle handlowe mają stabilny charakter lub ich sezonowość jest przewidywalna bo zależy od produktu bądź pory roku, to te informacje pozwolą sformułować trafne przewidywania dotyczące wielkości przyszłych przychodów. Natomiast jeśli użytkownik wielokrotnie wraca do strony internetowej przed dokonaniem konwersji, warto zastanowić się nad stworzeniem projektu strony internetowej, który zapewniłby szybszą i łatwiejszą drogę do tej konwersji.

Typowe uruchomienie modułu odpowiedzialnego za śledzenie transakcji w sklepach internetowych będzie miało miejsce po zakończeniu procesu zakupu. Najczęściej realizowane jest to na stronie potwierdzającej dokonanie zakupu. Podczas wyświetlenia strony zawierającej potwierdzenie zakupu, dodatkowo w tle, w sposób niewidoczny dla użytkownika i z wykorzystaniem języka JavaScript zachodzi zapisanie informacji o całej transakcji. W przypadku usługi Google Analytics ten proces wygląda następująco:

⁵ D. Chaffey, M. Patron. 2012. From web analytics to digital marketing optimization: Increasing the commercial value of digital analytics. *Journal of Direct, Data and Digital Marketing Practice* 14, 30-45.

1. Uruchomienie modułu śledzenia transakcji e-commerce.
2. Dodanie transakcji.
3. Dodanie przedmiotów do transakcji.
4. Wysłanie danych.
5. Skasowanie danych.

Moduł transakcji zwykle jest ładowany do przeglądarki internetowej niezależnie od głównego modułu analityki internetowej i tylko wtedy, gdy jest wymagany, aby zmniejszyć rozmiar głównej biblioteki. Po jego załadowaniu, będą dostępne nowe polecenia. Po załadowaniu modułu śledzenia transakcji, zostanie utworzony obiekt transakcji.

W przypadku utworzenia transakcji w usłudze Google Analytics, będzie ona realizowana za pomocą następującego kodu:

```
ga('ecommerce:addTransaction', {  
  'id': '1234',  
  'affiliation': 'Acme Clothing',  
  'revenue': '11.99',  
  'shipping': '5',  
  'tax': '1.29'  
});
```

Transakcja zawiera dane o jej numerze, przychodzie, wysokości podatku oraz koszcie przesyłki. Następnie do transakcji dodawane są towary. Dodanie towaru realizowane jest poprzez kolejny kod:

```
ga('ecommerce:addItem', {  
  'id': '1234',  
  'name': 'Fluffy Pink Bunnies',  
  'sku': 'DD23444',  
  'category': 'Party Toys',  
  'price': '11.99',  
  'quantity': '1'  
});
```

Produkt zawiera dane o numerze transakcji do której należy, nazwę produktu, numer SKU, kategorię produktu, cenę oraz liczbę sztuk. Po dodaniu wszystkich produktów do transakcji jest ona wysyłana do centralnego serwera zbierającego dane.

Ten moduł śledzenia transakcji w sklepach ponadto można uzupełnić informacją o walucie. Przekazując tylko wartość transakcji, system analityki internetowej będzie je przypisywał wszystkie do domyślnej waluty. Jeśli natomiast sklep realizuje sprzedaż w wielu walutach, to wraz z transakcją można przesłać informację o walucie zapisaną w standardzie ISO 4217.

Znane z wielu systemów analityki internetowej śledzenie transakcji oparte jest głównie o dane zebrane i przesłane ze strony z potwierdzeniem zakupu. Obecnie może to być zbyt mało informacji, które mogą zostać wykorzystane do podniesienia liczby konwersji w sklepie internetowym. Badanie zachowania użytkownika wyłącznie przez to co włożył do koszyka a następnie zakupił to korzystanie z niewielkiej ilości dostępnych danych o tym, czego użytkownik w sklepie dokonał. Niedawno wprowadzony moduł ulepszony śledzenia e-

commerce w usłudze Google Analytics pozwala zbierać znacznie więcej danych niż do tej pory było to możliwe⁶.

Ulepszone śledzenie e-commerce

Największa różnica pomiędzy tradycyjnym śledzeniem konwersji w sklepie internetowym a ulepszonym e-commerce polega na mierzeniu większej liczby zdarzeń. Tradycyjne śledzenie e-commerce mierzy jedno zdarzenie - wyświetlenie danych o transakcji, dodanie do niej produktów i przesłanie ich do centralnego serwera zbierającego dane. Zwykle dzieje się to na stronie z potwierdzeniem zakupu.

Natomiast moduł ulepszanego śledzenia e-commerce mierzy 10 różnych zdarzeń. Do tych zdarzeń należą:

1. Wyświetlenie produktu na liście produktów
2. Kliknięcie produktu na liście produktów
3. Wyświetlenie produktu na stronie produktu
4. Dodanie produktu do koszyka
5. Usunięcie produktu z koszyka
6. Wyświetlenie wewnętrznych banerów promocyjnych
7. Kliknięcie wewnętrznych banerów promocyjnych
8. Lejek sprzedażowy
9. Sprzedaż koszyka
10. Zwrot towaru

Wprowadzenie śledzenia tych dodatkowych zdarzeń pozwala uzyskać wiele nowych informacji o zachowaniach zakupowych w sklepach internetowych. Nawet gdy w sklepie są oferowane najlepsze produkty i oferty promocyjne, to projekt i użyteczność sklepu muszą być dostosowane do użytkowników, tak aby zapewniać łatwość przeglądania produktów, znajdowania specyfikacji, dodawania towarów do koszyków i finalizowania transakcji.

Mierzenie dodatkowych zdarzeń w zakresie zachowań zakupowych i nabywczych daje wgląd w to, jak użytkownicy korzystają z zawartości sklepu internetowego poprzez wyświetlanie produktów, dodawanie i usuwanie produktów z koszyka, rozpoczynanie, porzucanie i finalizowanie transakcji.

Mierzenie dodatkowych zdarzeń w zakresie skuteczności ekonomicznej pozwala zbierać informacje o przychodach i wskaźnikach konwersji, które generują produkty, średnią liczbę produktów znajdujących się w transakcji, średnią wartość zamówienia, zrealizowane zwroty oraz stawki, przy jakich użytkownicy dodają produkty do koszyka i dokonują zakupów po przejrzaniu stron z informacjami o produkcie.

Domyślne wprowadzenie monitorowanych zdarzeń dostarcza dane dla indywidualnych produktów, można to rozszerzyć o zbieranie danych o kategorii lub producencie, by także ocenić informacje z ich perspektywy.

Mierzenie dodatkowych zdarzeń w zakresie promocji produktu, pozwala zbierać dane o wewnętrznych i zewnętrznych działaniach marketingowych. Zewnętrzne działania promocyjne jak śledzenie kodu partnerskiego pozwalają mierzyć przychody, transakcje i średnią wartość zamówienie, które są powiązane z witrynami partnerskimi, przyciągającymi użytkowników do sklepu internetowego. Można śledzić te same dane dla kuponu na

⁶ R. Bose, (2009), "Advanced analytics: opportunities and challenges", Industrial Management & Data Systems, Vol. 109 Iss 2 ss. 155 – 172.

poziomie zamówienia, bądź korzystać z kuponu dla produktu. Pomiar korzystania z kuponów dla produktów pozwala sprawdzić skuteczność tego kuponu pod kątem przychodu i dokonanych unikalnych zakupów.

Mierzenie dodatkowych zdarzeń w zakresie promocji wewnętrznych, np. wewnętrznych bannerów, które promują sprzedaż w innej części sklepu internetowego, pozwala śledzić wyświetlenia, kliknięcia oraz współczynnik klikalności dla tych promocji.

Mierzenie dodatkowych zdarzeń w zakresie przeglądania list z produktami pozwala sprawdzić, które listy produktów wyświetliły się użytkownikom, które produkty znalazły się w tych listach oraz które z nich zostały kliknięte, wyświetlone i jaki był współczynnik klikalności dla każdej listy, pozycji na liście i produktu.

Wdrożenie modułu ulepszonych śledzenie e-commerce wymaga korzystania z najnowszej wersji modułu analityki internetowej oferowanej w tym systemie, uruchomieniu modułu ulepszonych e-commerce, na tej samej zasadzie zgodnie z którą należy uruchomić tradycyjne śledzenie e-commerce w sklepie internetowym oraz przekazywania wszystkich danych o zdarzeniach do centralnego serwera danych.

Przekazywanie danych można realizować w oparciu o komendy systemu analityki internetowej, podobne swoją strukturą do tradycyjnego śledzenie e-commerce lub wykorzystać pośrednią warstwę danych tzw. datalayer. Datalayer to fragment skryptu JavaScript zapisany w kodzie strony internetowej, którego zawartość może zostać odczytana przez narzędzie pośrednie np. Google Tag Manager.

Wszystkie dodatkowe dane zebrane przy pomocy ulepszonych modułu śledzenia e-commerce można wykorzystać do stworzenia analizy popytu i sprzedaży. Proces takiej analizy mógłby wyglądać następująco. Dla wszystkich produktów w sklepie wyliczony zostaje indeks popytu. Indeks popytu tworzą takie dane, jak odsłony stron produktowych, dodania do koszyka, dodanie do „przechowalni”, czyli wszystkie parametry, które wskazują na zainteresowanie danym produktem. Następnie dla wszystkich produktów w sklepie wyliczony zostaje indeks sprzedaży, czyli przychód wygenerowany z danego produktu. Zestawiając indeks popytu z indeksem sprzedaży, można przygotować prostą macierz, składającą się z 4 pól:

1. Wysoki popyt, wysoka sprzedaż.
2. Wysoki popyt, niska sprzedaż.
3. Niski popyt, wysoka sprzedaż.
4. Nisko popyt, niska sprzedaż.

Produkty znajdujące się w pierwszym polu, to te którym można podnieść ceny i zwiększyć marżę.

Produkty, które znajdują się w drugim polu powinny mieć ulepszoną, bardziej zachęcającą ofertę.

Produkty, które znajdują się w trzecim polu potrzebują lepszego wyeksponowania w sklepie.

Produkty, które znajdują się w czwartym polu potrzebują lepszej ekspozycji w sklepie, ale poprawić należy także samą ofertę.

Spam w danych analitycznych

Od końca 2014 roku w wielu systemach analityki internetowej, między innymi Piwik, Google Analytics i innych można zaobserwować dane o charakterze spamu. Spam w rozumieniu analityki internetowej to zbiór danych, które nie mają pokrycia w rzeczywistym ruchu na stronie internetowej.

Spam występuje pod dwoma postaciami

1. Ruch z kanału odesłań z wizyt internetowych zawiera informacje o robotach internetowych, które odwiedziły stronę internetową, jednak był to automatyczny program odczytujący zawartość strony i przetwarzających znaczniki języka JavaScript.
2. Różne raporty zawierają fałszywe dane na temat zebranych informacji. Pojawiają się one tam przez import tych danych z wykorzystaniem zewnętrznych punktów styku. Na przykład w Google Analytics jest dostępny Measurement Protocol, który poprzez odwołania HTTP przesyła surowe dane do centralnego serwera Google Analytics. W ramach tych danych można przesłać fałszywe informacje o:
 - a. źródłach ruchu - na przykład nieprawdziwe informacje o odesłaniach, bądź ruchu bezpośrednim bez odesłań lub ruchu z wyszukiwarki internetowej,
 - b. wykorzystanych słowach kluczowych, które użytkownik wprowadził do wyszukiwarki internetowej a następnie przeszedł do strony internetowej,
 - c. transakcjach wykonanych w sklepie internetowym,
 - d. zdarzeniach zarejestrowanych na stronie internetowej.

Pierwszy rodzaj spamu można zablokować uruchamiając blokady na ruch z adresów internetowych, które znajdują się na globalnych, powszechnie znanych listach spamu ruchu z odesłań. Blokada ruchu w drugim przypadku jest niemożliwa, ponieważ ten ruch fizycznie nie istnieje. Do centralnego serwera danych dostarczane są fałszywe dane z pominięciem strony internetowej.

Jaki jest cel wprowadzania fałszywych danych do systemów analityki internetowej? Chodzi o procedurę wstrzykiwania plików cookie z informacją o partnerskim hiperłączu. Mechanizm wstrzykiwania fałszywych danych do systemów analityki internetowej wykorzystuje metodę socjotechniki, ludzką ciekawość. Użytkownik analizujący informacje z systemu analityki internetowej, gdy zobaczy nowe adresy odesłań z których pochodził ruch internetowy stronie internetowej jest ciekawy, co to za strony. Otwierając po kolei adresy tych stron, które znajdują się w systemie analityki internetowej, docelowo trafi nie na strony, których adresy wpisał, natomiast jest przekierowany pod adresy, zawierające odsyłacze partnerskie jednego z programów partnerskich (np. Amazon lub AliExpress).

Gdy użytkownik odwiedzi specjalnie przygotowany odsyłacz partnerski partnera, w jego przeglądarce zapisuje się plik cookie. Po powrocie na tę stronę w określonym terminie i dokonaniu zakupu np. w sklepie Amazon, prowizja od dokonanej transakcji zostanie zapisana na rzecz użytkownika, który przygotował odsyłacz partnerski dostępny pod adresem widocznym w systemie analityki internetowej. Plik cookie przechowuje informacje o tym, komu będzie się należała prowizja za zakup w sklepie internetowym.

Jak rozpoznać spam w danych zbieranych w systemach analityki internetowej? Pierwszą oznaką spamu będą nazwy domen, znajdujące się w źródle ruchu odesłania, które zachęcają do wykonania określonej akcji lub imitują forum internetowe. Przykładami takich adresów są:

- semalt.com
- darodar.com

- econom.co
- iloveitaly.com
- 7makemoneyonline.com
- saveyoutubevideo.com
- cclicking.com
- buttons-for-website.com
- tasteidea.com
- kambasoft.com

Rozpoznanie tych adresów jako fałszywych polega na zwróceniu uwagi na wzorec wejść na stronę internetową, który wygląda mniej więcej tak, że każde odwiedzinę na stronie to nowy użytkownik, który ma 100% współczynnik odrzuceń, widzi tylko jedną stronę na sesję oraz spędza 00:00:00 czasu na stronie. Istnieje wtedy wysokie prawdopodobieństwo, że to fałszywy ruch.

Na rysunku nr 5 znajduje się przykładowy raport zawierające informacje o ruchu z odesłań, który w całości zawiera fałszywe informacje.

Wymiar podstawowy: Źródło Strona wejścia Inne ▾

Wiersze wykresu Wymiar dodatkowy ▾ Sortowanie: Domyślne ▾

| Zródło ? | Pozyskiwanie | | | Zachowanie | | |
|--|---------------------------------------|---|---------------------------------------|--|---|--|
| | Sesje ? | % nowych sesji ? | Nowi użytkownicy ? | Współczynnik odrzuceń ? ↓ | Strony / sesja ? | Śr. czas trwania sesji ? |
| | 1 319 % całości: 65,85% (2 003) | 99,47% Śr. dla widoku danych: 95,66% (3,99%) | 1 312 % całości: 68,48% (1 916) | 79,83% Śr. dla widoku danych: 86,72% (-7,94%) | 1,11 Śr. dla widoku danych: 1,08 (3,64%) | 00:00:40 Śr. dla widoku danych: 00:00:26 (51,86%) |
| <input type="checkbox"/> 1. chinese-amazon.com | 13 (0,99%) | 100,00% | 13 (0,99%) | 100,00% | 1,00 | 00:00:00 |
| <input type="checkbox"/> 2. e-buycasy.com | 5 (0,38%) | 100,00% | 5 (0,38%) | 100,00% | 1,00 | 00:00:00 |
| <input type="checkbox"/> 3. erot.co | 1 (0,08%) | 0,00% | 0 (0,00%) | 100,00% | 1,00 | 00:00:00 |
| <input type="checkbox"/> 4. free-social-buttons.com | 19 (1,44%) | 100,00% | 19 (1,45%) | 100,00% | 1,00 | 00:00:00 |
| <input type="checkbox"/> 5. hongfanji.com | 3 (0,23%) | 33,33% | 1 (0,08%) | 100,00% | 1,00 | 00:00:00 |
| <input type="checkbox"/> 6. qualitymarketzone.com | 25 (1,90%) | 100,00% | 25 (1,91%) | 100,00% | 1,00 | 00:00:00 |
| <input type="checkbox"/> 7. site10.free-floating-buttons.com | 4 (0,30%) | 100,00% | 4 (0,30%) | 100,00% | 1,00 | 00:00:00 |
| <input type="checkbox"/> 8. site2.floating-share-buttons.com | 39 (2,96%) | 100,00% | 39 (2,97%) | 100,00% | 1,00 | 00:00:00 |

Rysunek nr 5, Fałszywe dane w systemie analityki internetowej

Źródło: Google Analytics, opracowanie własne

Bardzo podobnie będzie wyglądać raport z modułów udostępniających dane na temat transakcji w sklepie internetowym, zarejestrowanych zdarzeń na stronie internetowej lub wykorzystanych słowa kluczowych, które zostały wpisane do wyszukiwarki internetowej. Niektóre dane mogą być fałszywe, wprowadzone do systemu analityki internetowej w celu sfalszowania ruchu na stronie i skierowaniu użytkownika przeglądającego te dane do partnerskich odsyłaczy.

Drugi rodzaj spamu w danych analitycznych może zostać wykluczony tylko na podstawie konfiguracji narzędzia do analityki internetowej. Taka konfiguracja opiera się na zestawie filtrów wykluczających ruch z fałszywych odesłań, wykluczający ruch z urządzeń które nie posiadają podanej rozdzielczości ekranu, dołączających ruch wyłącznie pochodzący z określonego hosta, zazwyczaj ten samej nazwy jaką jest strona internetowa. Można stosować

jeszcze bardziej precyzyjne metody filtrowania, jednak każda z nich opiera się na integracji systemu analityki internetowej ze stroną internetową.

Podsumowanie

Systemy analityki internetowej dostarczają wielu informacji osobom, które na ich podstawie mogą podjąć trafne decyzje. Będą to decyzje w zakresie prowadzenia handlu internetowego, wykorzystywania informacji o zaangażowaniu użytkownika w stronę internetową oraz efektywności źródeł ruchu z których przychodzą użytkownicy. System analityki internetowej wydają się być niezbędnym narzędziem, które dostarcza informacji o stronie internetowej i jej przydatności.

Bibliografia

1. Ranjit Bose, (2009), "Advanced analytics: opportunities and challenges", *Industrial Management & Data Systems*, Vol. 109 Iss 2 ss. 155 – 172.
2. Dave Chaffey, Mark Patron. 2012. From web analytics to digital marketing optimization: Increasing the commercial value of digital analytics. *Journal of Direct, Data and Digital Marketing Practice* 14, 30-45.
3. Avinash Kaushik, (2010) *Web Analytics 2.0. Świadome rozwijanie witryn internetowych*, Helion, Gliwice.
4. Kazuo Nakatani Ta-Tao Chuang, (2011), "A web analytics tool selection method: an analytical hierarchy process approach", *Internet Research*, Vol. 21 Iss 2 ss. 171 – 186
5. H. Pakkala, K. Presser, T. Christensen, (2012) Using Google Analytics to measure visitor statistics: The case of food composition websites. *International Journal of Information Management* 32, 504-512.
6. Beatriz Plaza, (2009), "Monitoring web traffic source effectiveness with Google Analytics", *Aslib Proceedings*, Vol. 61 Iss 5 ss. 474 - 482