Monitorowanie Aplikacji Czytelnika systemu dLibra z wykorzystaniem Google Analytics

Treść tego dokumentu jest pisana z myślą o osobach, które mają już podstawowe pojęcie o działaniu i możliwościach Google Analytics. Osoby, które nie znają tego narzędzia powinny w pierwszej kolejności zapoznać się z materiałami dostępnymi bezpłatnie w Akademii Analytics prowadzonej przez Google.

Spis treści

- Spis treści
- Instalacja kodu śledzącego Google Analytics
- Zaawansowana konfiguracja Google Analytics
  - Usunięcie nieistotnych parametrów z adresów WWW
  - Konfiguracja śledzenia wyszukiwania wewnątrz biblioteki cyfrowej
  - Konfiguracja celów
    - Użytkownik w ramach sesji użył wyszukiwania prostego
    - Użytkownik w ramach sesji użył wyszukiwania zaawansowanego
    - Użytkownik w ramach sesji użył filtrowania wyników wyszukiwania
    - Użytkownik w ramach sesji użył sortowania wyników wyszukiwania
    - Użytkownik w ramach sesji przeglądał indeksy wartości metadanych
    - Użytkownik w ramach sesji wyświetlił zawartość obiektu cyfrowego
    - Użytkownik w ramach sesji pozostał przynajmniej pięć minut na stronie z treścią obiektu cyfrowego
    - Użytkownik w ramach sesji pobrał obiekt cyfrowy w postaci archiwum ZIP
    - Użytkownik w ramach sesji spędził na stronie WWW biblioteki cyfrowej łącznie przynajmniej pięć minut
- Przykłady analizy danych
  - Popularność obiektów - liczba wyświetleń strony z metadanymi i wyświetleń on-line
  - Popularność obiektów - liczba wyświetleń on-line i pobrań jako archiwum ZIP
  - Zainteresowanie obiektami - czas spędzany na stronach z treścią obiektów

Instalacja kodu śledzącego Google Analytics

Aby rozpocząć korzystanie z Google Analytics (dalej: GA), należy w pierwszej kolejności zainstalować kod śledzący GA w stronach generowanych przez Aplikację Czytelnika. Kod ten można znaleźć w panelu administracyjnym Google Analytics, w sekcji Usługa / Informacje o śledzeniu / Kod śledzenia, na przykład:

blocked URL

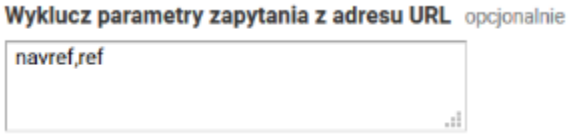
Cały kod (od `<script>` do `</script>`) należy umieścić w pliku *WEB-INF/components/templates/AnalyticsComponent.vm*. Po zapisaniu zmian w pliku, kod śledzący powinien być od razu poprawnie osadzany. Efektem tego będzie zarówno widoczność kodu śledzącego w kodzie HTML stron WWW Aplikacji Czytelnika, jak i rozpoczęcia raportowania danych do GA.

Zaawansowana konfiguracja Google Analytics

Poza opisanym powyżej zainstalowaniem kodu śledzącego zalecane jest również skonfigurowanie zaawansowanych funkcji Google Analytics. Najistotniejsze zaawansowane ustawienia opisano poniżej.

Usunięcie nieistotnych parametrów z adresów WWW

Konfigurowanie usunięcia z danych analitycznych nieistotnych w tym kontekście parametrów z adresów WWW możliwe jest w opcjach administracyjnych GA, w sekcji Widok danych, w Ustawieniach widoku. Należy tam wpisać w pole Wyklucz parametry zapytania z adresu URL wartość: navref,ref

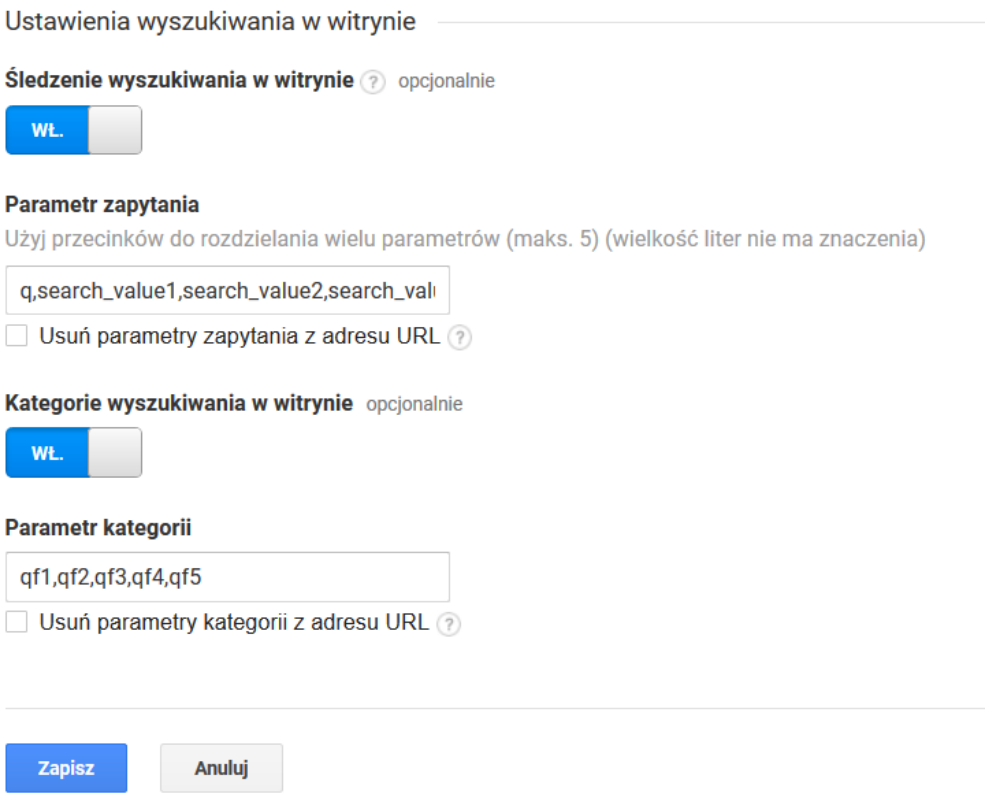


Konfiguracja śledzenia wyszukiwania wewnątrz biblioteki cyfrowej

Konfigurowanie śledzenia wyszukiwania wewnątrz biblioteki cyfrowej możliwe jest w opcjach administracyjnych GA, w sekcji Widok danych, w Ustawieniach widoku. Należy tam włączyć (patrz zrzut ekranu poniżej) Śledzenie wyszukiwania w witrynie, a także włączyć Kategorie wyszukiwania w witrynie oraz wpisać:

- W pole Parametr zapytania ciąg: q,search_value1,q,search_value2,q,search_value3,q,search_value4 - spowoduje to śledzenie zapytań prostych oraz zapytań zaawansowanych (pierwszych czterech elementów składowych).
- W pole Parametr kategorii ciąg: qf1,qf2,qf3,qf4,qf5 - spowoduje to śledzenie do pięciu równoczesnych filtrów wyszukiwania stosowanych przez użytkowników.

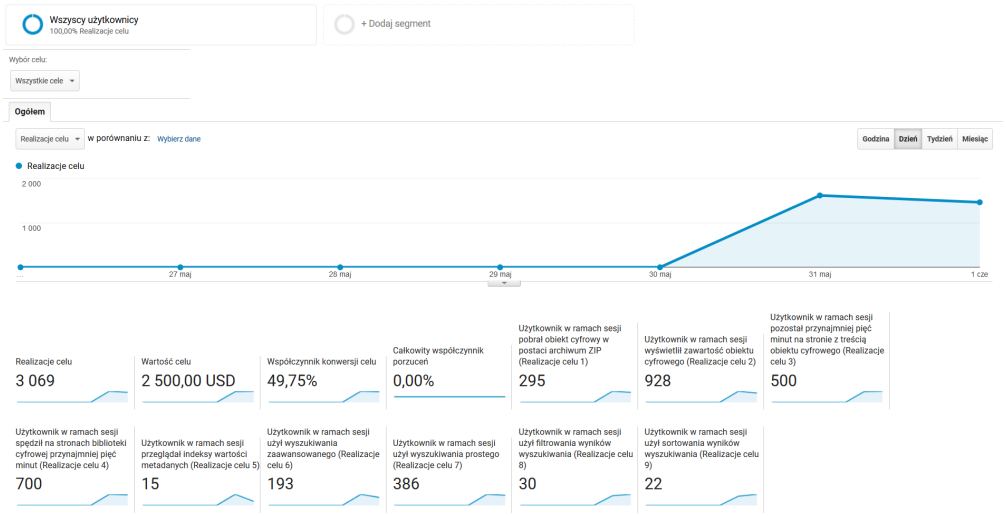
Raporty z wyszukiwania w witrynie można znaleźć w części analitycznej GA, w sekcji ZACHOWANIE / Wyszukiwane hasła.



Konfiguracja celów

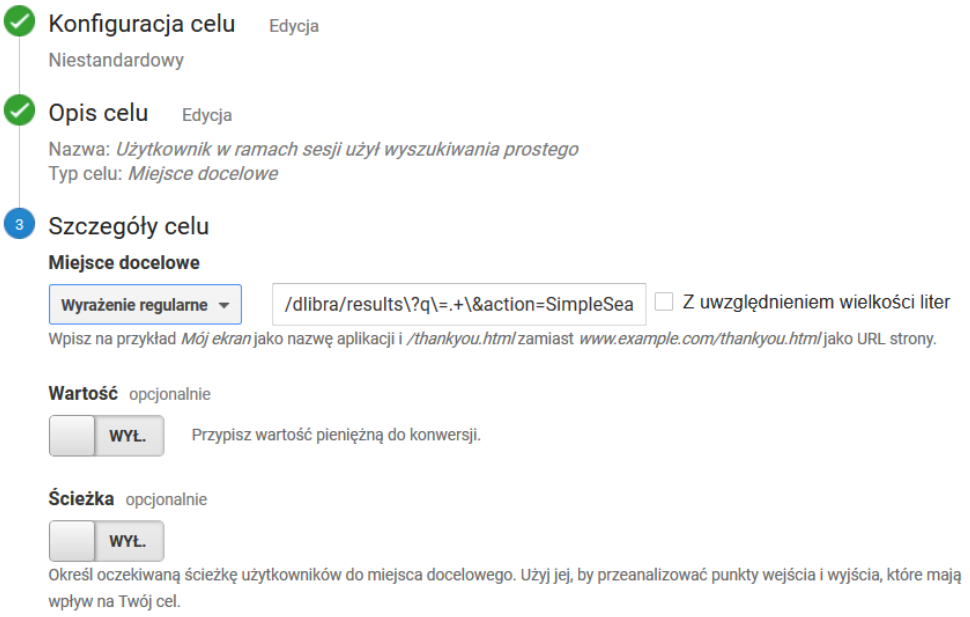
Cele w GA służą do monitorowania tego, w jakiej liczbie sesji użytkowników ma miejsce określone, istotne z naszego punktu widzenia, działanie. W przypadku sklepu internetowego takim działaniem może być dodanie przedmiotu do koszyka, dokonanie zakupu czy wysłanie formularza kontaktowego. W kontekście biblioteki cyfrowej proponujemy zdefiniowanie szeregu celów opisanych poniżej.

Cele definiuje się w części administracyjnej GA, w sekcji Widok danych / Cele. Po zdefiniowaniu celów rozpoczyna się w GA gromadzenie danych niezbędnych do wyświetlana statystyk dotyczących celów. Po dniu czy dwóch od zdefiniowania celów pierwsze efekty można sprawdzić w GA w części analitycznej, w sekcji KONWERSJE / Cele / Przegląd.



Użytkownik w ramach sesji użył wyszukiwania prostego

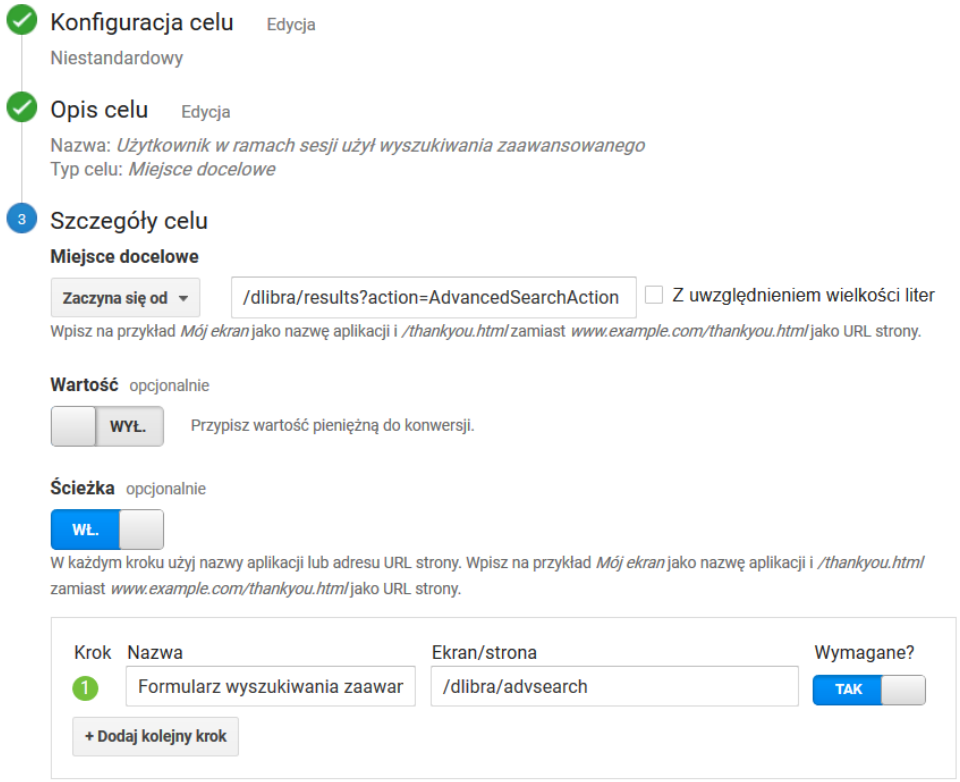
Ustawienia celu jak na zrzucie ekranu poniżej. Wartość wyrażenia regularnego: /dlibra/results\?q\=.+\&action=SimpleSearchAction\&.+



Użytkownik w ramach sesji użył wyszukiwania zaawansowanego

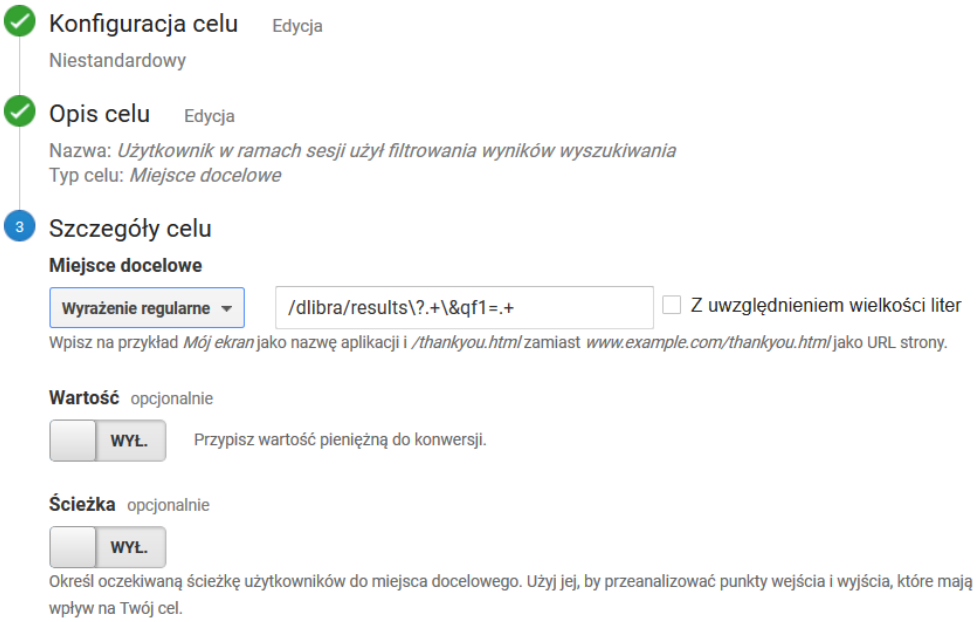
Ustawienia celu takie jak na zrzucie ekranu poniżej. Wartości poszczególnych pól:

- Wartość w polu "Zaczyna się od": /dlibra/results?action=AdvancedSearchAction
- Przełącznik "Ścieżka": WŁ.
- Wartości w sekcji "Krok 1":
    - Nazwa: Formularz wyszukiwania zaawansowanego
    - Ekran/strona: /dlibra/advsearch
    - Wymagane: TAK



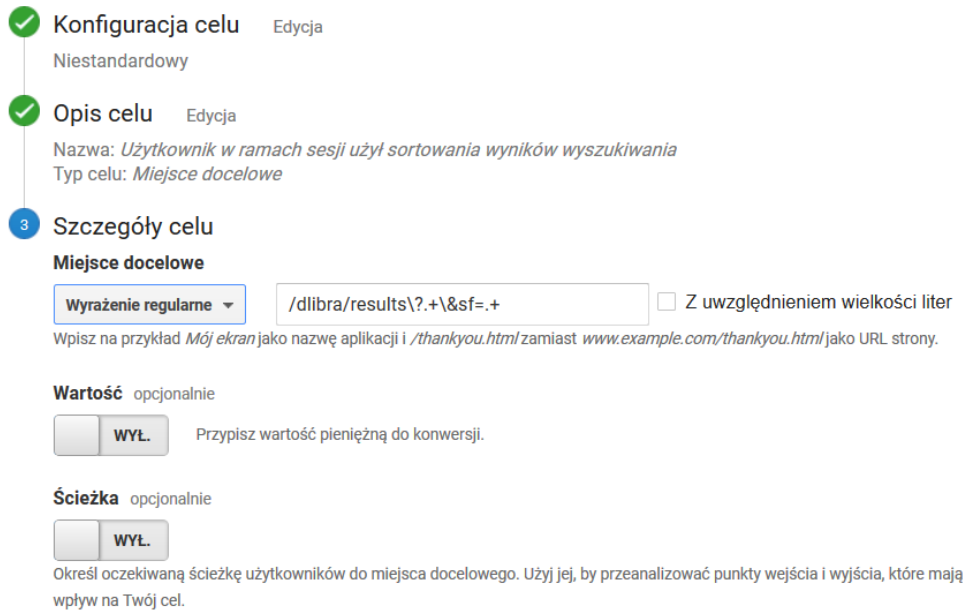
Użytkownik w ramach sesji użył filtrowania wyników wyszukiwania

Ustawienia celu jak na zrzucie ekranu poniżej. Wartość wyrażenia regularnego: /dlibra/results\?.+\&qf1=.+



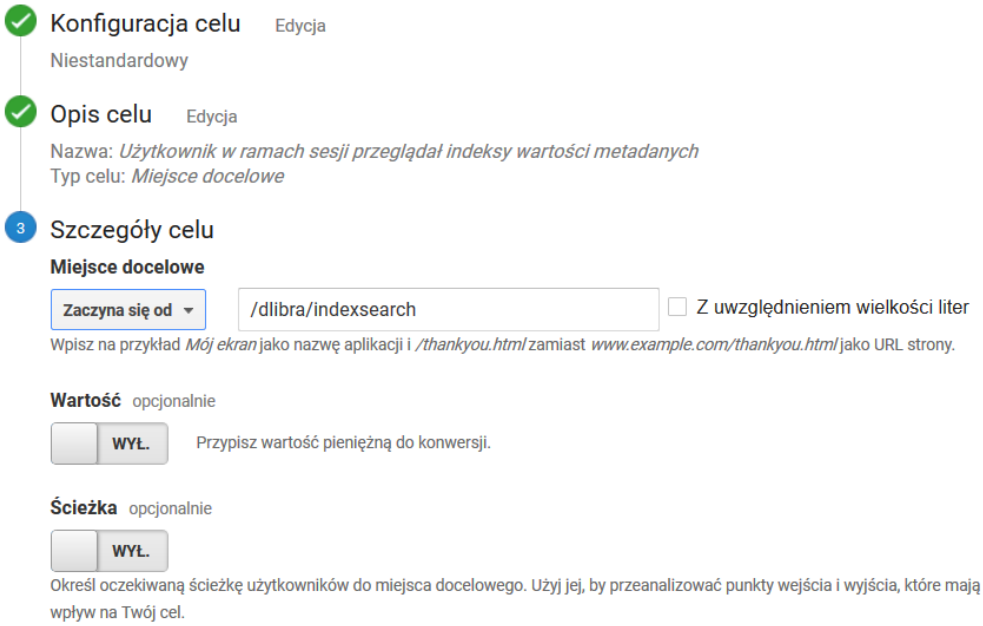
Użytkownik w ramach sesji użył sortowania wyników wyszukiwania

Ustawienia celu jak na zrzucie ekranu poniżej. Wartość wyrażenia regularnego: /dlibra/results\?.+\&sf=.+



Użytkownik w ramach sesji przeglądał indeksy wartości metadanych

Ustawienia celu takie jak na zrzucie ekranu poniżej. Wartość w polu Zaczyna się od: /dlibra/indexsearch



Użytkownik w ramach sesji wyświetlił zawartość obiektu cyfrowego

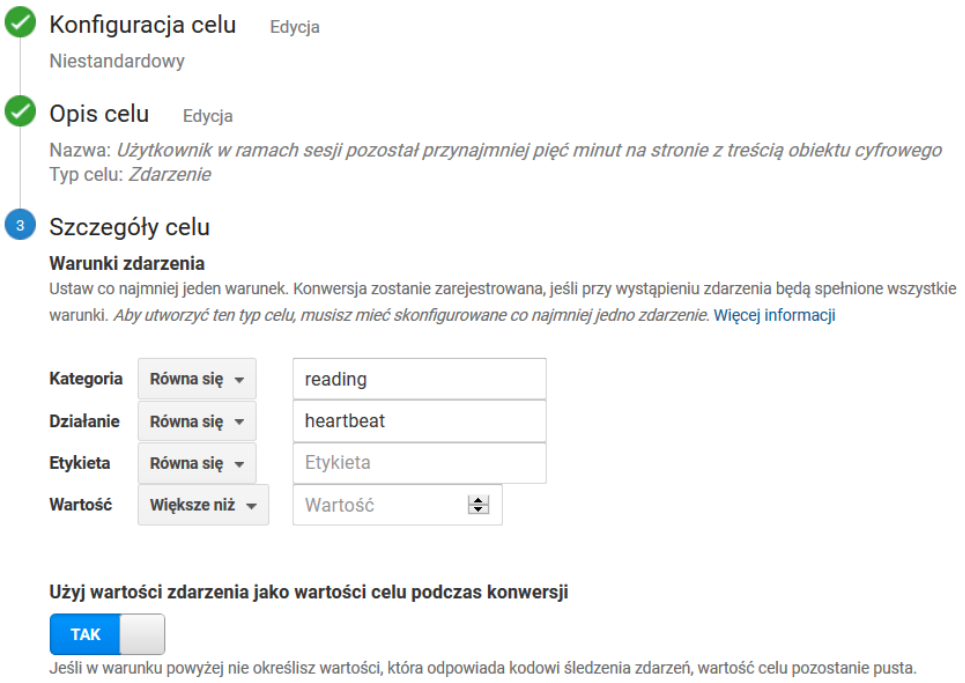
Ustawienia celu jak na zrzucie ekranu poniżej. Wartość wyrażenia regularnego: /dlibra/publication/[0-9]+/edition/[0-9]+/content.*

Użytkownik w ramach sesji pozostał przynajmniej pięć minut na stronie z treścią obiektu cyfrowego

Ustawienia celu jak na zrzucie ekranu poniżej. Wartości poszczególnych pól:

- Kategoria: reading
- Działanie: heartbeat
- Etykieta: *pozostaje puste*
- Wartość: *pozostaje puste*
- Użyj wartości zdarzenia jako wartości celu podczas konwersji: TAK

Monitorowanie czasu spędzonego na stronie z treścią obiektu cyfrowego jest realizowane poprzez dedykowane skrypty raportujące wbudowane w kod systemu dLibra, w stronę na której prezentowana jest treść obiektu. Skrypty te powodują, że gdy strona jest otwarta, co 5 minut do GA wysyłane jest zdarzenie (opisane jako [kategoria: reading, działanie: hearbeat]). Dzięki temu możliwe jest monitorowanie czasu jaki użytkownik spędza na interakcji z treścią obiektu (np. przegląda plik PDF). Interakcja taka, bez dodatkowego wsparcia po stronie systemu biblioteki cyfrowej, wymyka się skryptom śledzącym GA, powodując też znaczne zaniżanie pomiarów czasu trwania sesji użytkowników.

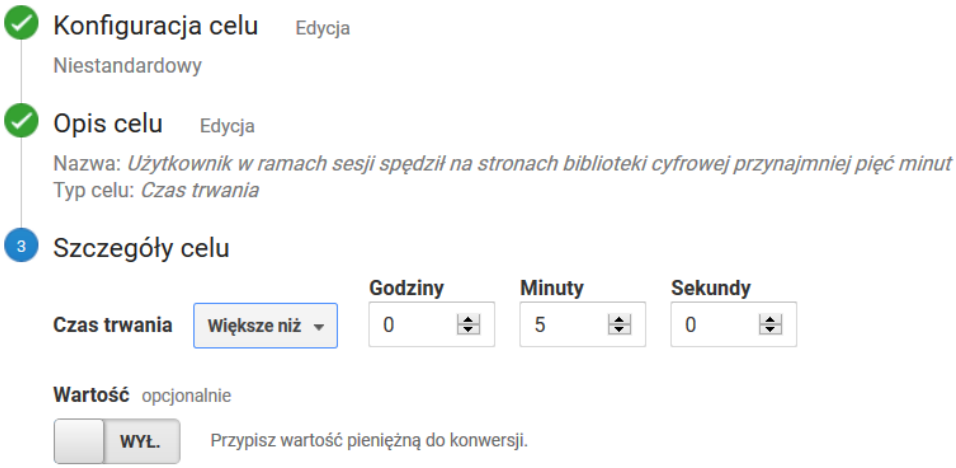


Użytkownik w ramach sesji pobrał obiekt cyfrowy w postaci archiwum ZIP

Ustawienia celu jak na zrzucie ekranu poniżej. Wartość wyrażenia regularnego: /Content/[0-9]+/zip/

Użytkownik w ramach sesji spędził na stronie WWW biblioteki cyfrowej łącznie przynajmniej pięć minut

Ustawienia celu jak na zrzucie ekranu poniżej.



Przykłady analizy danych

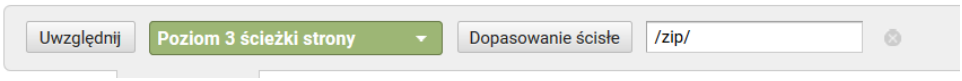
Poniżej przedstawiono przykłady analizy danych dotyczących ruchu użytkowników w bibliotece cyfrowej. Przykłady te absolutnie nie wyczerpują możliwości GA, jednak zostały dobrane w taki sposób, aby zwrócić uwagę na wybrane możliwości GA, ciekawe w kontekście bibliotek cyfrowych.

Popularność obiektów - liczba wyświetleń strony z metadanymi i wyświetleń on-line

1. Wybrać raport: ZACHOWANIE / Analiza treści
2. W tabeli, w kolumnie "Poziom 1 ścieżki strony" kliknąć /dlibra/
3. W tabeli, w kolumnie "Poziom 2 ścieżki strony" kliknąć /publication/
4. Widoczna jest tabela z identyfikatorami publikacji, posortowana domyślnie wg liczby odsłon, malejąco. Odsłona to w tym kontekście to:
    a. Dla publikacji z treścią: wyświetlenie strony z metadanymi lub wyświetlenie treści on-line.
    b. Dla publikacji bez treści (np. wyższego rzędu elementów publikacji grupowych - np. całe czasopismo): wyświetlenie strony z opisem /strukturą danego elementu publikacji grupowej.
5. Kliknięcie w identyfikator publikacji w kolumnie "Poziom 3 ścieżki strony" powoduje wyświetlenie tabeli dotyczącej sposobu interakcji z daną publikacją. W tabeli dla publikacji z treścią powinny znajdować się dwa wiersze:
    a. /zip/ - oznacza pobranie publikacji jako archiwum ZIP
    b. /<nazwa pliku głównego> - oznacza wyświetlenie publikacji on-line
6. Na kroku (3) powyżej można wprowadzić dodatkowe filtrowanie aby wyświetlić tylko pobrania. W tym celu należy:
    a. Dodać "Wymiar dodatkowy" (przycisk nad tabelą) o nazwie "Poziom 3 ścieżki strony"
    b. Włączyć zaawansowane filtrowanie wierszy tabeli (link "zaawansowane" po prawej stronie nad tabelą) i wybrać:

Popularność obiektów - liczba wyświetleń on-line i pobrań jako archiwum ZIP

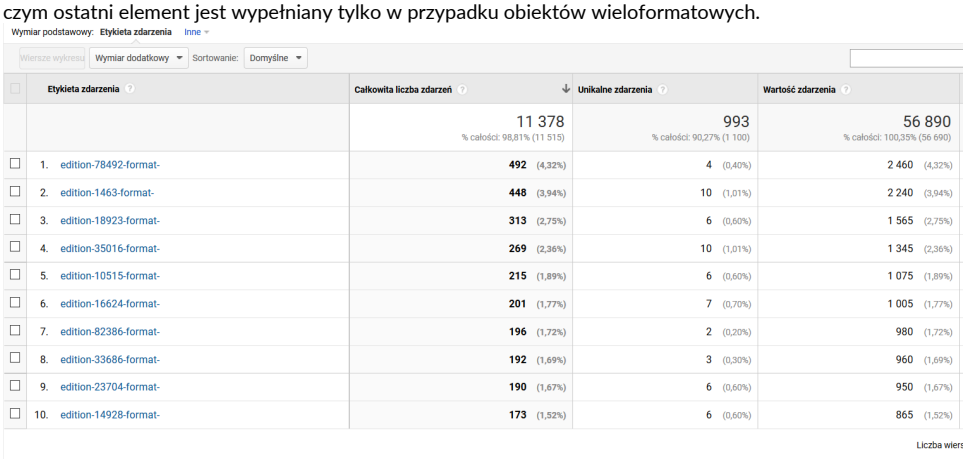
1. Wybrać raport: ZACHOWANIE / Analiza treści
2. W tabeli, w kolumnie "Poziom 1 ścieżki strony" kliknąć /Content/
3. Widoczna jest tabela z identyfikatorami publikacji, posortowana domyślnie wg liczby odsłon, malejąco. Odsłona to w tym kontekście liczba wyświetleń on-line i pobrań jako archiwum ZIP.
4. Kliknięcie w identyfikator publikacji w kolumnie "Poziom 2 ścieżki strony" powoduje wyświetlenie tabeli dotyczącej sposobu interakcji z daną publikacją. W tabeli powinny znajdować się dwa wiersze:
    a. /zip/ - oznacza pobranie publikacji jako archiwum ZIP
    b. /<nazwa pliku głównego> - oznacza wyświetlenie publikacji on-line
5. Na kroku (3) powyżej można wprowadzić dodatkowe filtrowanie aby wyświetlić tylko pobrania. W tym celu należy:
    a. Dodać "Wymiar dodatkowy" (przycisk nad tabelą) o nazwie "Poziom 3 ścieżki strony"
    b. Włączyć zaawansowane filtrowanie wierszy tabeli (link "zaawansowane" po prawej stronie nad tabelą) i wybrać:



Zainteresowanie obiektami - czas spędzany na stronach z treścią obiektów

1. Wybrać raport ZACHOWANIE / Zdarzenia / Najczęstsze zdarzenia.
2. W tabeli ze zdarzeniami kliknąć w "reading" (w kolumnie Kategoria zdarzenia).
3. W tabeli ze zdarzeniami kliknąć w "heartbeat" (w kolumnie Akcja zdarzenia).

4. Widoczna tabela przedstawia w kolumnie "Etykieta zdarzenia" listę identyfikatorów wydań, przy treści których użytkownicy spędzili najwięcej czasu. Każda z etykiet ma następującą strukturę: edition-<numeryczny identyfikator wydania>-format-<numeryczny identyfikator formatu>, przy czym ostatni element jest wypełniany tylko w przypadku obiektów wieloformatowych.


Wymiar podstawowy: **Etykieta zdarzenia** Inne

Wiersze wykresu | Wymiar dodatkowy | Sortowanie: Domyślne

	Etykieta zdarzenia	Całkowita liczba zdarzeń ↓	Unikalne zdarzenia	Wartość zdarzenia
		11 378 % całości: 98,81% (11 515)	993 % całości: 90,27% (1 100)	56 890 % całości: 100,35% (56 690)
1.	edition-78492-format-	**492** (4,32%)	4 (0,40%)	2 460 (4,32%)
2.	edition-1463-format-	**448** (3,94%)	10 (1,01%)	2 240 (3,94%)
3.	edition-18923-format-	**313** (2,75%)	6 (0,60%)	1 565 (2,75%)
4.	edition-35016-format-	**269** (2,36%)	10 (1,01%)	1 345 (2,36%)
5.	edition-10515-format-	**215** (1,89%)	6 (0,60%)	1 075 (1,89%)
6.	edition-16624-format-	**201** (1,77%)	7 (0,70%)	1 005 (1,77%)
7.	edition-82386-format-	**196** (1,72%)	2 (0,20%)	980 (1,72%)
8.	edition-33686-format-	**192** (1,69%)	3 (0,30%)	960 (1,69%)
9.	edition-23704-format-	**190** (1,67%)	6 (0,60%)	950 (1,67%)
10.	edition-14928-format-	**173** (1,52%)	6 (0,60%)	865 (1,52%)

Liczba wiers

5. Interpretacja poszczególnych kolumn jest następująca:
    a. Całkowita liczba zdarzeń - łączna liczba pięciominutowych interwałów spędzonych przez użytkowników nad treścią danego wydania
    b. Unikalne zdarzenia - liczba unikalnych sesji użytkowników, w których użytkownik spędził nad treścią danego wydania przynajmniej pięć minut
    c. Wartość zdarzenia - łączna liczba minut spędzona przez użytkowników nad treścią danego wydania
6. Dla ułatwienia analizy danych można dodać wymiar dodatkowy do tabeli (przycisk tuż nad tabelą). Sugerowane wymiary to:
    a. Strona - spowoduje dodanie kolumny z linkami do poszczególnych obiektów
    b. Tytuł strony - spowoduje dodanie kolumny z tytułami poszczególnych obiektów