

Analiza ruchu w aplikacji webowej

Projektowanie i programowanie systemów internetowych II

mgr inż. Krzysztof Rewak

11 stycznia 2020

Wydział Nauk Technicznych i Ekonomicznych

Państwowa Wyższa Szkoła Zawodowa im. Witelona w Legnicy

Plan prezentacji

1. Klasyczna analityka
2. Przydatna analityka
3. Straszna analityka
4. Podsumowanie

Vinaora Visitors Counter

012462

	Today	11
	Yesterday	33
	This week	554
	Last week	751
	This month	1791
	Last month	3674
	All	12462

Online now: 2

Your IP: 127.0.0.1

FIREFOX 3.0.8, WINDOWS

Now: 2009-04-17 20:03

Visitors Counter 1.6

Czasy, w których takie liczniki gości można było znaleźć bardzo często, słusznie już minęły.

Ludzie wykształcili od tego czasu wiele przydatnych narzędzi.
Przynajmniej część z nich warto znać.

Jak to ugryźć?

Gdzie można podpiąć nasłuchiwanie ruchu w naszej aplikacji?

Praktycznie rzecz biorąc wszędzie: zarówno na webowym backendzie, webowym frontendzie czy aplikacji mobilnej.

Co ugryźć?

Czego można dowiedzieć się podczas nasłuchiwania ruchu w naszej aplikacji?

Wszystkiego. W granicach rozsądku. Zazwyczaj.

Klasyczna analityka

Badanie zapytań HTTP

Większość frameworków oferuje proste sposoby na wydobycie informacji o IP klienta.

Przykładowo:

```
public function returnIP(Request $request): JsonResponse {  
    return response()->json(["ip" => $request->ip()]);  
}
```


A znając IP można zbadać już naprawdę wiele ciekawych rzeczy.

← → ⓘ ip-api.com/json/186.27.58.141

JSON Nieprzetworzone dane Nagłówki

Zapisz Kopiuuj Zwiń wszystkie Rozwiń wszystkie

```
as: "AS28024 Nuevatel PCS de Bolivia S.A."
city: "Villa Imperial de Potosí"
country: "Bolivia"
countryCode: "BO"
isp: "Nuevatel PCS de Bolivia S.A."
lat: -19.5892
lon: -65.7509
org: "NuevatelV"
query: "186.27.58.141"
region: "P"
regionName: "Potosí"
status: "success"
timezone: "America/La_Paz"
zip: ""
```

Logowanie zdarzeń

Warto rozważyć logowanie ruchu na stronie wewnątrz własnej aplikacji.
Jest to proste do wykonania, ale warto pamiętać o kilku szczegółach.

Logowanie zdarzeń

2019-01-09 16:16:06	37.47.130.112	● INFO	Courses list requested and delivered.
2019-01-09 16:13:31	37.47.130.112	● INFO	Authentication status requested and delivered.
2019-01-09 13:30:49	156.17.194.43	● INFO	Grades requested for group { id: 16 } by student { no: ██████ } and delivered.
2019-01-09 13:30:49	156.17.194.43	● INFO	Groups requested for course { id: 7 } and delivered.
2019-01-09 13:30:49	156.17.194.43	● INFO	Courses requested for semester { id: 3 } and delivered.
2019-01-09 13:30:49	156.17.194.43	● INFO	Semesters requested and delivered.
2019-01-09 13:30:46	156.17.194.43	● INFO	Courses requested for semester { id: 3 } and delivered.
2019-01-09 13:30:46	156.17.194.43	● INFO	Groups requested for course { id: 7 } and delivered.
2019-01-09 13:30:46	156.17.194.43	● INFO	Semesters requested and delivered.
2019-01-09 13:30:42	156.17.194.43	● INFO	Groups requested for course { id: 7 } and delivered.
2019-01-09 13:30:42	156.17.194.43	● INFO	Semesters requested and delivered.
2019-01-09 13:30:42	156.17.194.43	● INFO	Courses requested for semester { id: 3 } and delivered.
2019-01-09 13:30:42	156.17.194.43	● INFO	Groups requested for course { id: 7 } and delivered.
2019-01-09 13:30:39	156.17.194.43	● INFO	Courses requested for semester { id: 3 } and delivered.
2019-01-09 13:30:39	156.17.194.43	● INFO	Semesters requested and delivered.
2019-01-09 13:30:39	156.17.194.43	● INFO	Courses requested for semester { id: 3 } and delivered.
2019-01-09 13:30:36	156.17.194.43	● INFO	Semesters requested and delivered.
2019-01-09 13:30:31	156.17.194.43	● INFO	Authentication status requested and delivered.
2019-01-09 12:05:31	77.255.141.192	● INFO	Courses requested for semester { id: 3 } and delivered.

Logowanie zdarzeń

- serwery HTTP (Apache, nginx) najczęściej robią to już za nas w postaci access- i errorlogów;
- warto zastanowić się czy lepiej logować dane do pliku czy do bazy danych;
- jeszcze bardziej warto się zastanowić po jakim czasie dane należy usuwać;

Istnieje nawet standard RFC 5424, który opisuje poziomy logów:

- 100 DEBUG
- 200 INFO
- 250 NOTICE
- 300 WARNING
- 400 ERROR
- 500 CRITICAL
- 550 ALERT
- 600 EMERGENCY

Co można zrobić z takimi logami?

- zbadać zerwane linki;
- sprawdzić czy ktoś nie próbował się włamać do zastrzeżonych miejsc;
- sprawdzić popularność produktów;
- zbadać w którym miejscu procesu zakupowego klient najczęściej rezygnuje;
- i setki innych zastosowań...

Przydatna analityka

Królem webowej analityki od lat jest **Google Analytics**.

Stosowane na frontendzie, Google Analytics jest bardzo proste w użyciu. W bazowej wersji wymaga jedynie załączenia odpowiedniego skryptu, a następnie wywołania jednej metody `ga()`.

Google Analytics (a!ć!)

```
<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',
'https://www.google-analytics.com/analytics.js','ga');

ga('create', 'UA-4815162342-1', 'auto');
ga('send', 'pageview');
</script>
```

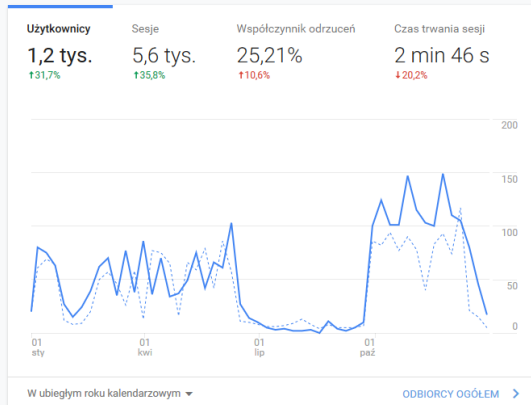
Oczywiście każdy sensowniejszy javascriptowy framework już od dawna ma własne wrappery, które niestraszają zminifikowanym kodem.

Przykładowa implementacja GA w Vue.js:

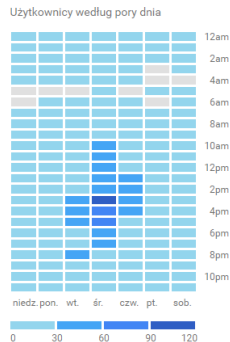
```
Vue.use(VueAnalytics, {  
  id: "UA-4815162342-1",  
  checkDuplicatedScript: true,  
  router  
})
```

Google Analytics

Strona główna Google Analytics



Kiedy użytkownicy odwiedzają Twoją witrynę lub aplikację?



Ostatnie 90 dni ▾

Google Analytics

Użytkownicy

2 450



Nowi użytkownicy

2 467



Sesje

11 670



Sesje na użytkownika

4,76



Odsłony

55 803



Strony/sesja

4,78



Śr. czas trwania sesji

00:03:10

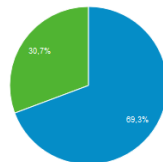


Współczynnik odrzuceń

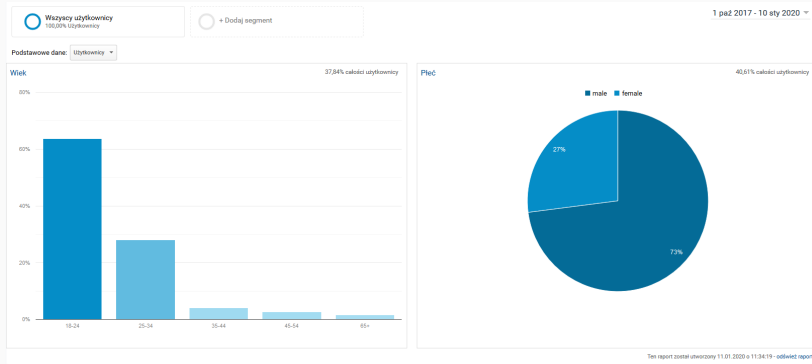
23,92%



■ New Visitor ■ Returning Visitor



Google Analytics



Google Analytics

Podobne kategorie (zasięg)

44,94% całości użytkowników

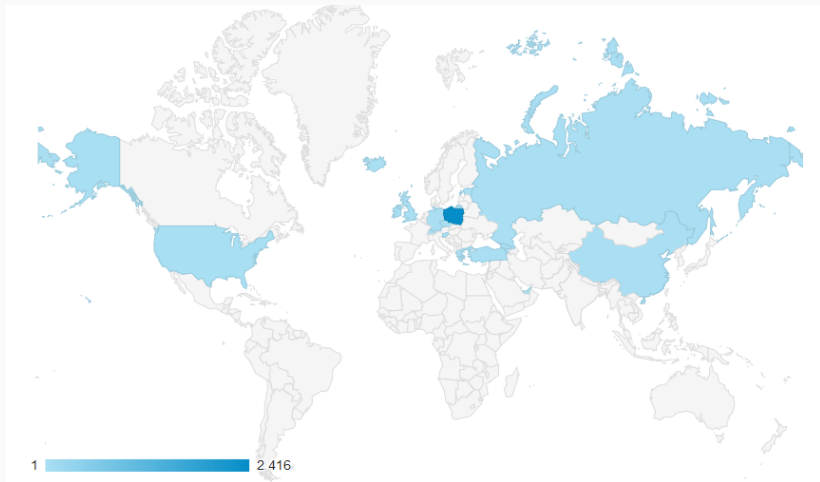
4,05%	<div></div>	Technology/Technophiles
3,91%	<div></div>	Media & Entertainment/Gamers
3,48%	<div></div>	Media & Entertainment/Movie Lovers
3,46%	<div></div>	Shoppers/Value Shoppers
3,03%	<div></div>	Media & Entertainment/Music Lovers
2,89%	<div></div>	Food & Dining/Fast Food Cravers
2,72%	<div></div>	Lifestyles & Hobbies/Shutterbugs
2,64%	<div></div>	Sports & Fitness/Health & Fitness Buffs
2,58%	<div></div>	Media & Entertainment/Gamers/Hardcore Gamers
2,56%	<div></div>	Sports & Fitness/Sports Fans

Inna kategoria

45,14% całości użytkowników

4,44%	<div></div>	Arts & Entertainment/Humor/Funny Pictures & Videos
3,23%	<div></div>	Arts & Entertainment/Celebrities & Entertainment News
2,87%	<div></div>	Science/Mathematics
2,63%	<div></div>	Games/Computer & Video Games/Browser Games
2,52%	<div></div>	Reference/General Reference/Dictionaries & Encyclopedias
2,43%	<div></div>	Arts & Entertainment/TV & Video/Online Video
2,31%	<div></div>	Internet & Telecom/Email & Messaging
2,12%	<div></div>	Sports/Team Sports/Soccer
2,05%	<div></div>	Games/Roleplaying Games
2,05%	<div></div>	Online Communities/Photo & Video Sharing/Photo & Image Sharing

Google Analytics



Google Analytics

<input type="checkbox"/>	System operacyjny ?	Pozyskiwanie		
		Użytkownicy ? ↓	Nowi użytkownicy ?	Sesje ?
		2 450 % całości: 100,00% (2 450)	2 469 % całości: 100,08% (2 467)	11 670 % całości: 100,00% (11 670)
<input type="checkbox"/>	1. Windows	1 668 (67,59%)	1 682 (68,12%)	7 656 (65,60%)
<input type="checkbox"/>	2. Android	504 (20,42%)	504 (20,41%)	3 316 (28,41%)
<input type="checkbox"/>	3. iOS	223 (9,04%)	225 (9,11%)	448 (3,84%)
<input type="checkbox"/>	4. Linux	49 (1,99%)	34 (1,38%)	120 (1,03%)
<input type="checkbox"/>	5. Macintosh	23 (0,93%)	23 (0,93%)	129 (1,11%)
<input type="checkbox"/>	6. (not set)	1 (0,04%)	1 (0,04%)	1 (0,01%)

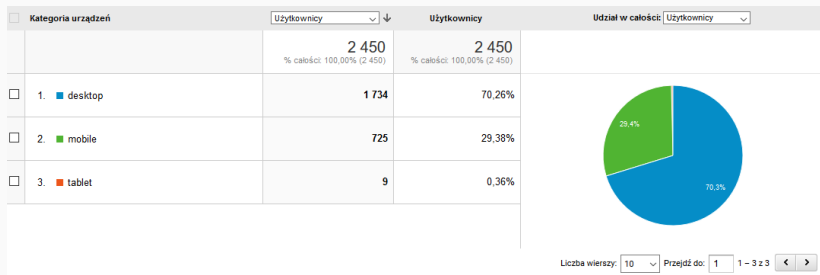
Google Analytics

<input type="checkbox"/>	Przeglądarka ?	Pozyskiwanie		
		Użytkownicy ? ↓	Nowi użytkownicy ?	Sesje ?
		2 450 % całości: 100,00% (2 450)	2 469 % całości: 100,08% (2 467)	11 670 % całości: 100,00% (11 670)
<input type="checkbox"/>	1. Chrome	1 492 (60,85%)	1 503 (60,87%)	7 953 (68,15%)
<input type="checkbox"/>	2. Firefox	521 (21,25%)	526 (21,30%)	2 217 (19,00%)
<input type="checkbox"/>	3. Safari	222 (9,05%)	223 (9,03%)	425 (3,64%)
<input type="checkbox"/>	4. Opera	83 (3,38%)	82 (3,32%)	799 (6,85%)
<input type="checkbox"/>	5. Android Webview	57 (2,32%)	57 (2,31%)	83 (0,71%)
<input type="checkbox"/>	6. Edge	40 (1,63%)	40 (1,62%)	97 (0,83%)





















Google Analytics

<input type="checkbox"/>	Wersja dodatku Flash [?]	Pozyskiwanie		
		Użytkownicy [?] ↓	Nowi użytkownicy [?]	Sesje [?]
		2 450 % całości: 100,00% (2 450)	2 469 % całości: 100,08% (2 467)	11 670 % całości: 100,00% (11 670)
<input type="checkbox"/>	1. (not set)	2 132 (84,74%)	2 122 (85,95%)	10 235 (87,70%)
<input type="checkbox"/>	2. 32.0 r0	113 (4,49%)	101 (4,09%)	615 (5,27%)
<input type="checkbox"/>	3. 29.0 r0	89 (3,54%)	82 (3,32%)	228 (1,95%)
<input type="checkbox"/>	4. 27.0 r0	82 (3,26%)	77 (3,12%)	225 (1,93%)
<input type="checkbox"/>	5. 28.0 r0	40 (1,59%)	34 (1,38%)	103 (0,88%)
<input type="checkbox"/>	6. 31.0 r0	36 (1,43%)	31 (1,26%)	152 (1,30%)

Google Analytics



Google Analytics

Strona		Odsłony	% Odsłony
1. /kursy		9 716	 17,41%
2. /		9 642	 17,28%
3. /oceny		5 310	 9,52%
4. /kursy/1		4 855	 8,70%
5. /aktualnosci		2 234	 4,00%
6. /kursy/3		1 231	 2,21%
7. /kontakt		1 025	 1,84%
8. /kursy/4		898	 1,61%
9. /faq		833	 1,49%
10. /zarzadzaj		832	 1,49%

[wyświetl pełny raport](#)

Google Analytics



Straszna analityka

Istnieje cała masa innych rozwiązań.

Popularnym może być New Relic (raczej wykorzystywany po stronie backendu), Kissmetrics (do śledzenia konkretnych użytkowników) czy CrazyEgg (do generowania heatmap).

Należy pamiętać, że granica prywatności to rzecz bardzo względna i może się okazać, że aplikacja śledzi naprawdę każdy ruch użytkownika.

Jakim problemem (od strony technicznej) jest sprawdzenie w jakie miejsca na stronie klika się najczęściej?

Jakim problemem (od strony technicznej) jest zrobienie zrzutu ekranu użytkownika?

Jakim problemem (od strony moralnej) jest nagranie całej sesji użytkownika?

Czas na eksperyment!

Podsumowanie

Bibliografia i ciekawe źródła



<https://tools.ietf.org/html/rfc5424>

Pytania?

Kod prezentacji dostępny jest w repozytorium git pod adresem
<https://bitbucket.org/krewak/pwsz-ppsi2>



Wszystkie informacje dot. kursu dostępne są pod adresem
<http://pwsz.rewak.pl/kursy/6>

