

# Porównywanie stron internetowych

Lingaro sp. z.o.o

autor:  
Marcin Grzabka

Warszawa 21.03.2014

## Spis treści

1. Cel i założenia.....	3
1.1 Założenia:.....	3
1.2 Metoda porównywania.....	4
1.3 Metody porównywania obrazu.....	5
2. Uzyskane wyniki.....	6
3. Zagadnienia.....	7
3.1 Określenie błędu.....	7
3.2 Wyróżnienie elementu strony jako obiektu istotnego z punktu widzenia porównania wyglądu strony.....	8
3.3 Zagłębianie się wewnątrz hierarchii DOM.....	8
3.4 Realizowalność techniki podświetlania elementów.....	8
3.5 Renderowanie stron.....	9
3.5.1 Domyślnej wielkości renderowania .....	9
3.5.2 Uzyskanie screenshotu ekranu.....	9

## 1. Cel i założenia

Celem projektu jest porównywanie renderowania stron internetowych pomiędzy przeglądarkami oraz interpretacja uzyskanych wyników.

Poprzez porównywanie rozumie się:

- sprawdzenie różnic generowania elementów (napisów, ramek itp)
  - kolorystyki,
  - czcionek,
  - rozmiar
- sprawdzenie położenia elementów
  - przesunięcia,
  - przekrywanie

### 1.1 Założenia:

- porównaniu podlegają jedynie statyczne elementy strony (brak wsparcia dla rozwijanych menu, zakładek)
- rezultatem porównania jest obraz zawierający pierwotną witrynę wraz z nałożonymi różnicami (w postaci obrazu różnicowego)
- do generowanej strony konieczne jest dołączenie skryptu JavaScript
- porównywane są pojedyncze elementy (istnieje możliwość porównywania “poziomów zagłębień”)

## **1.2 Metoda porównywania**

Wypracowane podejście bazuje na porównywaniu odpowiadających sobie elementów, co pozwoliło uzyskać niewrażliwość na względne przesunięcia w obrazie. Poniżej przedstawiono wady i zalety rozwiązania.

### Zalety:

- Możliwość porównania dowolnego elementu
- Porównanie wybranych elementów pozwala na analizę różnic występujących tylko w tym obszarze
- Niewrażliwość na przesunięcia elementów (różnice w marginesach)

### Wady:

- Zagłębienia w hierarchii DOM powodują, że porównywane są obszary zawierające wcześniej sprawdzane elementy (a nawet się z nich składające).
- Czasochłonność generowania porównań możliwych zagłębień
- Niewrażliwość na przesunięcia elementów (różnice w marginesach)

Wyznaczanie elementów odpowiadających sobie, dokonywane jest przez użycie skryptu JavaScript (załącznik nr 1) według algorytmu:

1. ponumeruj w hierarchii DOM elementy strony uznane za istotne przy porównaniu.
2. rozpoczynając porównanie podświetl odpowiedni element (przyjęto kolor zielony, można dowolny, najlepiej nie występujący na stronie) a następnie wykonaj screenshot strony internetowej z podświetleniem.
3. na podstawie położenia elementu kolorystycznego z oryginalnej witryny wycinany jest odpowiedni element strony przeznaczony do porównania.
4. “zgaś” element oraz przejdź do punktu 2.

### 1.3 Metody porównywania obrazu

Dla wyodrębnionych elementów zgodnie z punktem 3 powyższego algorytmu liczone są momenty ciężkości poszczególnych składowych (R,G,B) według następującej zależności (w obu osiach):

$$x_c = \frac{\sum_{i=0}^n x_i v_i}{\sum_{i=0}^M v_i} \quad y_c = \frac{\sum_{i=0}^n y_i v_i}{\sum_{i=0}^M v_i}$$

gdzie:

$x_i$  - odległość od lewej krawędzi

$y_i$  - odległość od górnej krawędzi

$v_i$  - wartość intensywności pixela

Następnie obliczana jest odległość pomiędzy centrami odpowiednich kolorów obu obrazów z wykorzystaniem normy euklidesowej.

$$d_R = \sqrt{(x_R - \hat{x}_R)^2 + (y_R - \hat{y}_R)^2}$$

$$d_G = \sqrt{(x_G - \hat{x}_G)^2 + (y_G - \hat{y}_G)^2}$$

$$d_B = \sqrt{(x_B - \hat{x}_B)^2 + (y_B - \hat{y}_B)^2}$$

W przypadku gdy wartość któregoś ze współczynników odległości jest większa od 1.1 (dobrano doświadczalnie) wyliczany jest obraz różnicowy z obu porównywanych elementów.

## 2. Uzyskane wyniki

**Uwaga! Ta strona wykorzystuje pliki Cookie!** Szczegółowe informacje znajdziesz w polityce prywatności. Akceptuję politykę.

**Zarząd  
Transportu  
Miejskiego  
w Warszawie**

English | Deutsch | Русский

08 57  
11 marca 2014r.  
wtorek

szukaj w serwisie

strona główna

A A A

Fb Tw

wersja mobilna

mapa serwisu

kontakt

drukuj

<b>rozkłady jazdy</b>	<b>ostatnie zmiany</b>	<b>zaplanuj podróż</b>
<b>zaplanuj podróż</b>	<b>bus</b> dn. <b>12.03.2014</b> Zmiany w rejonie pl. Piłsudskiego oraz na Trakcie Królewskim w dniu 12 marca 2014	start
<b>informacje</b>	<b>bus</b> od <b>12.03.2014</b> do <b>16.03.2014</b> Zamknięcie dla ruchu skrzyżowania ulic Trakt Lubelski/Skalnicowa	koniec
<b>mapy i schematy</b>	<b>bus</b> od <b>11.03.2014</b> do <b>12.03.2014</b> Nocne prace w rejonie pętli DW. CENTRALNY	data <input type="text" value="2014-03-11"/>
<b>bilety i opłaty</b>	<b>skm/kolej</b> od <b>09.03.2014</b> <b>Remont dalekobieżnej linii średnicowej – zmiany w komunikacji kolejowej</b>	godzina <input type="text" value="07:28"/> <input checked="" type="radio"/> odjazd <input type="radio"/> przyjazd
<b>zakup biletu</b>	<b>bus</b> od <b>08.03.2014</b> Zmiana lokalizacji przystanku SADY ŻOLIBORSKIE 06	szukaj <input type="button" value="ZTM"/> <input type="button" value="jakdojade"/>
<b>przepisy</b>	<b>bus</b> od <b>07.03.2014</b> Czasowa zmiana lokalizacji przystanków GEN. MACZKA 01 i 02	 <b>OD 1.01.2014 R. NOWA TARYFA BILETOWA</b>
<b>karta warszawiaka</b>	<b>bus</b> od <b>07.03.2014</b> Czasowa zmiana trasy linii 187 w związku z remontem estakad Trasy Łazienkowskiej	 <b>KARTA WARSZAWIAKA KARTA MŁODEGO WARSZAWIAKA</b>
<b>osobista WKM</b>	<b>bus</b> od <b>03.03.2014</b> Czasowa zmiana tras linii 187, 194, 307, N35, N85 w rejonie ul. Orłąt Lwowskich	<b>newsletter</b>
<b>parkuj i jedź</b>	<b>bus</b> od <b>01.03.2014</b> Uruchomienie przystanków na trasach linii L-2 i 703	<b>Komunikacja jest rodzaju żeńskiego</b>
<b>lotnisko</b>	<b>bus</b> od <b>01.03.2014</b> Czasowa zmiana tras linii 119, 142, 305, 702 i N72 na ul. Trakt Lubelski	 Potraktujmy Dzień Kobiet jako okazję do przybliżenia sylwetek pięciu kobiet, których praca zawodowa jest związana z komunikacją miejską i które – w ramach swoich obowiązków
<b>serwis prasowy</b>	 <b>KARTA WARSZAWIAKA KARTA MŁODEGO WARSZAWIAKA</b>	
<b>inwestycje</b>	<b>aktualności</b>	<b>informacje prasowe</b>
<b>zamówienia publiczne</b>	 2 160	<b>Za 10 miesięcy dojedziemy tramwajem na Tarchomin</b>
<b>projekty europejskie</b>		28 lutego
<b>akcje społeczne</b>		
<b>hobby</b>		
<b>wynajem powierzchni</b>		
<b>praca w ZTM</b>		
<b>kontakt</b>		

rys 1. Przykład podświetlenia elementu strony www



rys 2. Wygenerowana strona z zaznaczonymi różnicami (obraz różnicowy)

### 3. Zagadnienia

Poniżej przedstawiono powstałe zagadnienia i problemy wraz z krótkim ich opisem.

#### 3.1 Określenie błędu

Należy sformułować określenie błędu w kontekście przetwarzania wybranego fragmentu strony.

Dla przykładu: wygenerowanie tekstu na stronie nieznacznie różniącego się wielkością czcionki, wygenerowanie innego rodzaju obramowania poprzez różne przeglądarki, przesunięcie położenia elementu itp.

Obecnie brane pod uwagę są tylko znaczne różnice kolorystyczne (których

efektem może być np. inna postać ramki, ale nie jest to bezpośrednio rozpoznawane),

### ***3.2 Wyróżnienie elementu strony jako obiektu istotnego z punktu widzenia porównania wyglądu strony.***

Istotą problemu jest określenie elementów ważnych z punktu widzenia wygenerowanego obrazu czyli zawierających jak największą ilość informacji (elementów leżących w ich wnętrzu).

Na chwilę obecną program podświetla i porównuje:

- DIV,
- TABLE,
- P,H1,H2,H3,H4,H5,H6,A,CENTER,STRONG

### ***3.3 Zagłębianie się wewnątrz hierarchii DOM***

W przypadku braku ograniczeń w maksymalnym dozwolonym poziomie zagłębienia, dochodzi do porównania elementu leżącego wewnątrz innego elementu, a następnie obiektu zewnętrznego (np. div w div).

W chwili ta sytuacja nie jest rozwiązana i porównywane są zagłębione w sobie elementy.

### ***3.4 Realizowalność techniki podświetlania elementów***

Nie jest wykazane, że istnieje możliwość podświetlenia (zmiany koloru tła) dowolnego elementu znajdującego się na stronie www.

Na chwilę obecną takie sytuacje są ignorowane i w przypadku braku odpowiadającego obrazu różnicowego nie następuje porównanie.



### **3.5 Renderowanie stron**

Podczas prac napotkano problemy dotyczące renderowania samej witryny w różnych przeglądarkach.

#### **3.5.1 Domyślnej wielkości renderowania**

Tylko w Chrome, istniała możliwość zmiany domyślnego rozmiaru (przeglądarki FireFox i Internet Explorer generowały strony w tym samym rozmiarze). Podejście polegające na przeskalowaniu strony/elementu powoduje wprowadzenie błędów do obrazu, które mimo braku zmian (w oryginalnej witrynie) wykażą nieprawidłowość w porównaniu.

#### **3.5.2 Uzyskanie screenshotu ekranu**

Udało się uzyskać jedynie screenshot strony w granicach jej widoczności (fragment witryny widoczny na ekranie). Próba rozwiązania problemu może być podejście polegające na uzyskaniu uchwytu do elementu okna zawierającego renderowaną stronę.

Załącznik nr 1 (Kod JS)