

Porównanie algorytmów wyszukiwania Google i Yandex

Oleh Bihun*, Marek Miłoś

Politechnika Lubelska, Instytut Informatyki, Nadbystrzycka 36B, 20-618 Lublin, Polska

Streszczenie. W artykule przedstawiono rezultaty analizy porównawczej wyszukiwarek Google i Yandex. Analizę przeprowadzono pod kątem szybkości wyszukiwania oraz liczby znalezionych wyników. Przy pomocy metody kryteriów ważonych porównano Google i Yandex dla wyszukiwania w języku angielskim i rosyjskim. Analiza wielokryterialna pozwoliła wskazać wyszukiwarke o szybszych i wydajniejszych algorytmach.

Słowa kluczowe: wyszukiwarka; porównanie; algorytmy wyszukiwania

* Autor do korespondencji.

Adres e-mail: oleh.bihun@pollub.edu.pl

Comparison Google's and Yandex's search algorithms

Oleh Bihun*, Marek Miłoś

^aInstitute of Computer Science, Lublin University of Technology, Nadbystrzycka 36B, 20-618 Lublin, Poland

Abstract. The article presents the results of a comparative analysis Google's and Yandex's search algorithms. The analysis was made from the point of view of search speed and number of results. Google and Yandex were compared for search in Russian and English language, using the method of weighted criteria. Multi-criteria analysis allowed to find faster and efficient search engine.

Keywords: search engine; comparison; search algorithms

*Corresponding author.

E-mail address: oleh.bihun@pollub.edu.pl

1. Wstęp

Na dzień dzisiejszy zasoby sieci Internet zawierają około 10 miliardów stron [1], do których ma dostęp każdy użytkownik sieci. Zatem dla znalezienia potrzebnej informacji należy korzystać z zaawansowanych i nowoczesnych wyszukiwarek.

W przypadku dużej liczby informacji pojawia się problem jej wyszukiwania i filtrowania. W dzisiejszych czasach jest już nadmiar danych i im więcej ich pojawia się, tym trudniej coś znaleźć [2].

A z drugiej strony, obecnie nie można sobie wyobrazić współczesnego życia bez Internetu i bez wyszukiwarek, za pomocą których można odnaleźć konkretne informacje w tej prawie nieskończonej objętości danych. Internet jest obecnie wszędzie – i w laptopach i telefonach, ale kluczem do poszukiwania są wyszukiwarki.

Aktualność problemu można rozważać w kategorii konfrontacji pomiędzy dużą liczbą informacji w sieci a słabymi możliwościami szybkiego jej wyszukania [3]. Algorytmy wyszukiwarek, które pozwalają szybko uszeregować dane i wyświetlić użytkownikom potrzebne wyniki - odgrywają tu bardzo ważną rolę. Algorytm wyszukiwarki jest główną częścią, która gwarantuje prawidłowość wyświetlanych danych.

W ramach pracy zostały przeanalizowane dwie popularne wyszukiwarki - Google i Yandex.

2. Metody oceny wyszukiwarek

Istnieje wiele różnych metod analizy wielokryterialnej, przy czym niektóre z nich zostały opracowane już wiele lat temu, a niektóre powstały całkiem niedawno. Wśród nich można wyróżnić podstawowe techniki, które są dziś powszechnie znane i stosowane w nauce.

Metoda kryteriów ważonych (ang. *Weighting Metod*) [4] polega na transformacji funkcji wielowymiarowej na jednowymiarową, czyli na połączeniu wielu funkcji w jedną [4]. Otrzymaną w ten sposób funkcję optymalizuje się za pomocą standardowych metod optymalizacji funkcji jednowymiarowej. Główną wadą tej metody jest problem z doborem odpowiedniej wartości wag dla poszczególnych kryteriów. Dobór ten w znaczący sposób wpływa na jakość otrzymanego rozwiązania.

Metoda optymalizacji hierarchicznej (ang. *Hierarchical Optimization Metod*) [4] polega na zastosowaniu optymalizacji jednowymiarowej krok po kroku, z wykorzystaniem poszczególnych kryteriów.

Pomysł, zastosowany w algorytmie VEGA (VEGA: Vector Evaluated Genetic Algorithm) [5], polega z kolei na podziale kategorii na podkategorie o tej samej liczbie elementów. Selekcja w obrębie każdej kategorii odbywa się niezależnie (jedna podkategoria jest odpowiedzialna za konkretne kryteria).

3. Obiekt badań

Do analizy porównawczej wybrano dwie popularne wyszukiwarki - Google i Yandex. Google jest popularna na całym świecie i posiada największy udział w rynku [6].

Yandex - to popularna wyszukiwarka w Rosji i bardzo znana wśród osób korzystających z języka rosyjskiego [7]. Yandex jest popularna w takich państwach jak: Rosja, Ukraina, Kazachstan, Białoruś. By opracować metodę porównywania wyszukiwarek, trzeba poznać ogólną zasadę ich działania.

Wyszukiwarka ma skomplikowaną strukturę, ale oprócz web-serwera czy bazy danych, główną częścią wyszukiwarek jest robot. Robot zajmuje się poszukiwaniem stron, pobieraniem ich do bazy danych i analizą [8]. Działa on za pomocą określonych algorytmów, które pozwalają mu analizować i szeregować dane [6]. To właśnie algorytmy odpowiedzialne za szeregowanie są obiektem badań [9]. Celem niniejszej pracy jest przeanalizowanie tych algorytmów oraz porównanie ich szybkości i wydajności.

4. Metoda badawcza

Główna teza pracy zawiera się w stwierdzeniu, że Google lepiej wyszukuje informacje w języku angielskim, a Yandex - w języku rosyjskim. Dlatego też zostały sformułowane dwie hipotezy badawcze:

H1: Google zwraca lepsze wyniki niż Yandex w języku angielskim.

H2: Yandex zwraca lepsze wyniki niż Google w języku rosyjskim.

Trzeba zauważyć, że trudno jest obiektywnie zmierzyć jakość wyszukiwania, czy jakość działania algorytmów [10]. Jest to związane przede wszystkim z niedostępnością danych o tych algorytmach – nie jest to bowiem informacja publiczna [11]. W związku z tym została opracowana metoda, która pozwoli na porównanie wyszukiwarek za pomocą konkretnych parametrów [12]. Trzeba zauważyć, że badania zostały przeprowadzone przy użyciu próbki losowej danych do zapytań, co pozwoliło w przejrzysty sposób ocenić i porównać systemy wyszukiwania.

Do porównania wybrano dwa kryteria, które zostały podzielone na dwie kategorie. W etapie finalnym wyszukiwarki są porównane za pomocą kryterium końcowego. Te kategorie to:

- 1) Jakość wyszukiwania w języku angielskim.
- 2) Jakość wyszukiwania w języku rosyjskim.

W etapie finalnym, kryterium końcowe powstaje jako suma powyższych. Wychodzi się bowiem z założenia, że każda z wyżej wymienionych kategorii w równym stopniu wpływa na ocenę końcową.

Analizując kierunek pracy podjęto decyzję o korzystaniu z uproszczonej metody ważonych kryteriów. Ocena przy pomocy metody ważonych kryteriów została podzielona na dwa etapy.

Na pierwszym etapie następuje przyznanie "wagi" dla określonych kryteriów lub kategorii. W ramach danej pracy można dołączyć określone wagi dla zadań, na przykład z różną częstotliwością. Celem pracy jest porównanie wyszukiwarek w aspekcie wyszukiwania dla każdego rodzaju zapytania, biorąc pod uwagę znaczenie każdego zapytania w zależności od użytkownika. Dlatego brano pod uwagę, że zapytania niskiej częstotliwości są tak samo ważne dla

użytkownika, jak i wysokiej częstotliwości. Zakłada się, że nawet rzadko zadawane pytania będą mieć znaczenie dla użytkownika. Na tym etapie przyznano jednakowe "wagi" dla wszystkich kategorii.

Na etapie drugim następuje obliczenie kompleksowych ocen. Ponieważ zostały przyjęte jednakowe "wagi" dla wszystkich kategorii, to obliczenia są realizowane według następującego wzoru [4]:

$$P = \sum_{k=1}^Q \left(\sum_{i=1}^N y_{ki} \frac{100}{N * m_k} \right) \quad (1)$$

gdzie:

- Q – liczba kategorii,
- N – liczba kryteriów w danej kategorii,
- y_{ki} – wartości i -tego kryterium w k -tej kategorii,
- m_k – maksymalna liczba punktów za kryterium.

5. Kryteria oceny wyszukiwarek

Kryteria oceny wyszukiwarek zostały dobrane tak, aby na podstawie uzyskanych w wyniku eksperymentu danych, można było w łatwy i obiektywny sposób je porównać.

Dla badania wyszukiwarek wykorzystano 500 najpopularniejszych słów kluczowych dla Google i Yandex. Przykładowe frazy prezentuje tabela 1.

Tabela 1. Frazy zastosowane do analizy wyszukiwarek

Frazy w języku angielskim	Frazy w języku rosyjskim
Hotels	Вк
Facebook	Авито
Facebook login	Фильмы онлайн
Amazon.com	Рамблер почта
Google earth	Яндекс почта
Scorpio	Русская весна
The hunsyellow pages	Телепрограмма
Map quest	Школьный портал

Ponieważ celem pracy było zbadanie aktualnych algorytmów, wykorzystano parametry, które są udostępniane przez same wyszukiwarki. To wyszukiwarka wyświetla użytkownikom czas i liczbę wyników. Te właśnie parametry stały się też kryteriami do oceny wyszukiwarek.

Jakość wyszukiwania w języku angielskim

Ta kategoria odpowiada za ocenę jakości wyszukiwania przy wprowadzaniu do wyszukiwarki słów kluczowych w języku angielskim. W celu oceny i opisu tej kategorii skorzystano z dwóch kryteriów:

- Czas wyszukiwania.
- Liczba zwracanych wyników.

Jakość wyszukiwania w języku rosyjskim

Ta kategoria odpowiada za ocenę jakości wyszukiwania przy wprowadzaniu słów kluczowych w języku rosyjskim. W celu oceny i opisu tej kategorii - skorzystano z tych samych kryteriów co w języku angielskim.

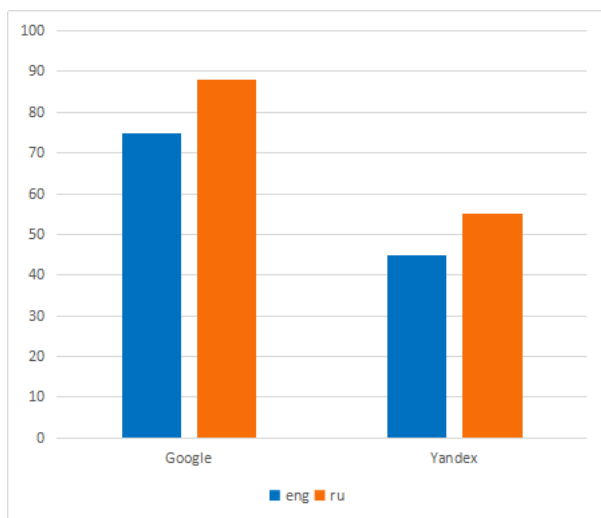
6. Rezultaty badań

Rezultaty badań można zaprezentować w postaci tabeli 2. Tabela prezentuje średni czas wyszukiwania t_{avg} oraz średnią liczbę zwracanych wyników n_{avg} dla wyszukiwania w języku angielskim (eng) i w języku rosyjskim (ru).

Tabela 2. Rezultaty badań wyszukiwarek

	Google		Yandex	
	eng	ru	eng	ru
$t_{avg}(sek)$	0.75	0.55	1.04	1.02
$n_{avg}(mln)$	326.6	572.9	40.8	171

Na podstawie tabeli 2 można stwierdzić, że Google ma lepsze parametry niż Yandex, co graficznie zaprezentowano w postaci diagramu – rys. 1.



Rys. 1. Porównanie Google i Yandex w kategoriach wyszukiwania w języku angielskim (eng) i języku rosyjskim (ru)

Google uzyskał lepszą ocenę niż Yandex w obydwu kategoriach. Rezultatem badań jest fakt, że wyszukiwarka Google zwraca więcej wyników w krótszym czasie podczas wyszukiwania zarówno w języku angielskim jak i rosyjskim.

Wynik końcowy prezentuje tabela 3.

Tabela 3. Ranking badanych wyszukiwarek

Pozycja	Wyszukiwarka	Kryterium końcowe
1	Google	81,5
2	Yandex	50

7. Wnioski

W ramach pracy przebadano i porównano parametry dla dwóch języków wyszukiwania, co pozwoliło sformułować wniosek, która wyszukiwarka zwraca lepsze wyniki i dla jakiego języka wyszukiwania.

Z otrzymanych wyników jednoznacznie można stwierdzić, że algorytmy wyszukiwarki Google są lepsze niż Yandex. Zatem, hipoteza H1 potwierdziła się, a hipoteza H2 – nie. Yandex zwraca średnio mniej wyników w dłuższym czasie.

Google korzysta z bardziej nowoczesnych i zaawansowanych algorytmów, niż Yandex, które pozwalają na szybsze wyszukiwanie informacji i zwracają większą liczbę wyników. Yandex jest w fazie ciągłego rozwoju,

również korzysta z nowoczesnych technologii, ale algorytmy przez nią stosowane, na chwilę obecną nie dorównują wyszukiwarce Google. Przyczyną tego mogą też być większe moce obliczeniowe Google (tj. lepsza warstwa techniczna wyszukiwarki). Dokonana analiza potwierdza tezę o bardzo zaawansowanych technologiach, stosowanych w Google.

W pracy zostały przeanalizowane tylko dwie wyszukiwarki. Ciekawym kierunkiem dalszych badań może być wskazanie innej metody oceny wyszukiwarek i ich algorytmów. Można też wybrać inne języki do porównania oraz inne wyszukiwarki, np. DuckDuckGo czy Ecosia.

Literatura

- [1] С. Симонович, Информатика. Базовый курс, СПб: «Питер» (2007), 85-88.
- [2] К. Маннинг, Введение в информационный поиск, Вильямс (2011), 115-117.
- [3] В. Белозеров, Эффективность систем Гугл и Яндекс для поиска учебного материала (2015), 254-255.
- [4] T. Borek, K. Cichosz, Wprowadzenie do optymalizacji wielokryterialnej, Akademia górniczo-hutnicza, Kraków, 2017
- [5] T. Lachowicz, Optymalizacja wielokryterialna decyzji w zagadnieniach bezpieczeństwa funkcjonowania podmiotu, Studia Ekonomiczne, Uniwersytet Ekonomiczny w Katowicach, no. 235 (2015), 144-145.
- [6] E. Schmidt, J. Rosenberg, How Google Works, 2014.
- [7] www.spravki.se-ua.net/yandex, Как работает яндекс [28.04.2017]
- [8] Z. Joyner, Slaying the Search Engine Dragon: Smarter, More Efficient Internet Searches, University of Georgia School of Law (2016), 11-15.
- [9] R. Meusel, P. Mika, R. Blanco, Focused Crawling for Structured Data, Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (2014), 1039-1048.
- [10] D. Nandy, H. Gohel, F. Kalyani, M. Khakhariya, Duckduckgo vs. Google search engine – comparative analysis, GRD Journals, Volume 2, Issue 1, 2016.
- [11] P. Chen, H. Xie, S. Maslov, S. Redner, Finding scientific gems with Google's PageRank algorithm, Journal of Informetrics, Boston, 2007