

Specyfikacja Wymagań

dla systemu

Trend Spotter

Wersja 1.2

Historia Wersji

Wersja	Data	Autor / Autorzy	Opis
1.0	12.03.2018	Marek Smolarczyk	Struktura dokumentu i podstawowe wymagania
1.1	19.03.2018	Miron Szewczyk	Specyfikacja wymagań
1.2	26.03.2018	Michał Pawłowski	Kilka poprawek

Zespół

1. Marek Smolarczyk (386007)
2. Miron Szewczyk (383504)
3. Michał Pawłowski (371303)
4. Bartek Wrona (386389)

Opis ogólny

System ma umożliwiać wykrywanie kształtujących się trendów i analizę historii popularności różnych tematów w serwisach internetowych. Z systemu będą mogli korzystać zarówno klienci indywidualni jak i firmy, i organizacje.

Na system składają się serwer WWW oraz crawler.

Założenia i zależności

1. System może gromadzić anonimowe informacje o sposobie użytkowania aplikacji.
2. System wymaga rejestracji w usłudze.
3. System gwarantuje prawidłowe działanie aplikacji na przeglądarkach:
 - Firefox (min. wersja .),
 - Chrome (min. wersja .).

Wymagania funkcjonalne

Aplikacja będzie umożliwiała korzystanie z przetworzonych danych zgromadzonych przez crawlera w następujące sposoby:

1. Wyświetlanie historii popularności wpisanych przez użytkownika wyrazów lub fraz odpowiadających zbiorom wyrazów.
2. Wyświetlanie najpopularniejszych zapytań użytkowników.
3. Wyświetlanie wyrazów, fraz szybko zyskujących na popularności (podczas wyliczania ogólnych statystyk pomijane są wyrazy z wcześniej przygotowanej listy, na przykład spójniki).

Dodatkowo użytkownik będzie miał możliwość:

4. Ograniczania zapytań do poszczególnych serwisów, poprzez wybór jednego lub wielu serwisów z rozwijanej listy.
5. Wyświetlenia wszystkich lub części komentarzy spełniających jego kryteria wyszukiwania.
6. Włączenia za pomocą checkboxa ignorowania polskich znaków w danym zapytaniu.
7. Włączenia za pomocą checkboxa ignorowania wielkich liter w danym zapytaniu.

Pozostałe wymagania:

8. Korzystanie z aplikacji wymaga zarejestrowania i zalogowania w serwisie. Możliwa jest rejestracja za pomocą kont Facebook i Gmail.
9. Dla każdego użytkownika jest przetrzymywana historia jego zapytań. Jest możliwe ponowne wykorzystanie dawnych zapytań poprzez wybranie ich z listy.

Wymagania niefunkcjonalne

1. Crawler obsługuje maksymalnie 10 serwisów. Dane zebrane przez crawler'a są przechowywane w bazie danych nie dłużej niż 6 miesięcy.
2. Dane przechowywane przez system są związane z użytkownikami, tak więc jest konieczne szyfrowanie.
3. System zapewnia obsługę pojedynczego zapytania użytkownika średnio w nie więcej niż 5 sekund.
4. System umożliwia jednoczesne płynne korzystanie z usługi przez nie więcej niż 10 użytkowników.
5. System zapewnia średni czas pomiędzy awariami nie krótszy niż 2000 godzin oraz średni czas naprawy około 12 godzin.
6. Przechowywane jest tylko pierwsze 8000 znaków każdego zebranego komentarza.
7. Baza danych jest odporna na ataki typu SQL-injection.
8. Każda strona jest crawlowana niezależnie. Żadna crawlowana strona nie zostanie od crawlera więcej niż jednego zapytania na sekundę.
9. Administrator aplikacji ma opcje wymuszenia na crawlerze przejrzienia całej strony. Czynność ta jest bardzo czasochłonna i może zająć do tygodnia.
10. Działanie crawlera opiera się na powracaniu pod wybrane linki z częstotliwością zależącą od dat zebranych pod nimi komentarzy. Im nowsza strona i im częściej pojawiają się na niej komentarze, tym częściej crawler tam wraca.
11. W bazie danych przechowywane są komentarze wraz z datą ich napisania, datą ich zaindeksowania przez crawlera i linkiem, pod jakim zostały znalezione.
12. Komentarze są przechowywane w bazie danych za pomocą kodowania UTF8.
13. Średni czas od powstania komentarza do jego zaindeksowania jest nie większy niż 24h.
14. Wymagania graficzne aplikacji (wraz z przykładami) opisane są w pliku