

Etap 2. Projekt Systemu Informatycznego

System do tworzenia rankingu stron dla
wybranych fraz w wyszukiwarce Google

Agnieszka Pustuła grupa Z615

1. Opis dziedziny problemu (wprowadzenie do tematu)

Podstawową rolą systemu jest monitorowanie pozycji wybranej strony w wyszukiwarce internetowej. Ponadto, program ma za zadanie śledzić ranking najpopularniejszych stron dla zadanej frazy oraz umożliwić generowanie raportów zawierających wykresy popularności.

Podstawowe zadania realizowane przez program:

- Wyszukiwanie najpopularniejszych stron dla danej frazy
- Generowanie raportów
- Ustawienie automatycznego raportowania danych o wynikach dla frazy w danych okresach czasowych w wyznaczonej przez użytkownika częstotliwości

Poszukiwanymi stronami, które byłyby oparciem dla projektowania możliwie najlepszej wersji systemu są dwa typy aplikacji webowych. Jeden z nich odwołuje się do aplikacji jako narzędzia typu "Crawler", zwane inaczej "Robotem internetowym". Robotami internetowymi nazywane są narzędzia będąca zautomatyzowanym programem zbierającym informacje na wybrany temat. Przykładem najpopularniejszego Crawlera jest Robot Indeksujący Google. Google porządkuje informacje na temat istniejących stron internetowych poprzez indeksację, czyli skatalogowanie witryn internetowych wraz z niezbędnymi danymi o rodzaju i treści strony.

Drugim typem poszukiwanej witryny pomocnej w wykonaniu systemu do tworzenia rankingu stron dla wybranych fraz jest aplikacja webowa skupiona na prezentacji danych. Jej głównym profilem jest przedstawianie zbiorów informacji w sposób przyjazny dla człowieka. Bez wątpienia najsukuteczniejszą formą prezentacji danych jest obraz.

2. Analiza stanu zastanego

Aplikacja internetowa wyszukiwarki Google

Aplikacja internetowa wyszukiwarki Google posiada mechanizmy, których wzorce warto wziąć pod uwagę przy projektowaniu części backendowej systemu do tworzenia rankingu.

Użytkownik wpisując frazę w wyszukiwarce internetowej Google oczekuje otrzymać zestaw najtrafniejszych wyników. Aby było to możliwe Google dużo wcześniej porządkuje informacje na temat istniejących stron internetowych poprzez indeksację. Zgodnie z oficjalnymi danymi Indeks Wyszukiwarki Google zawiera setki miliardów stron internetowych, a jego rozmiar znacznie przekracza 100.000.000 gigabajtów. Indeksacja stron dzieje się z użyciem zautomatyzowanych programów zbierających informację o strukturze, zawartości, a nawet kodzie witryny. Robot indeksujący krążąc po stronie zwraca szczególną uwagę na zmiany w istniejących już stronach, martwe linki (martwe linki czyli: usunięte podstrony, domeny które uległy wygaśnięciu, hiperłącza do niedziałającej strony) oraz linki prowadzące do nowych stron.

Szczególnie interesującym architektonicznie faktem jest reużywanie przez programy skanujące Google'a zebranych już wcześniej informacji na temat zindeksowanych stron. Tak skonstruowany algorytm zdecydowanie pozwala na ograniczenie zużycia zasobów obliczeniowych przy wykonywaniu powtarzalnych akcji i może być wykorzystany przy zbieraniu danych do projektu systemu rankingu stron.

Biorąc pod uwagę, że przygotowywany projekt systemu rankingu stron internetowych w wyszukiwarce Google będzie potrzebował jedynie informacji na temat pozycji strony po wpisaniu zadanego hasła, zbędne mu będą rozwiązania wchodzące w głąb struktury danej strony, a więc ta funkcjonalność robotów indeksujących Google jest zbędna.

Aplikacja internetowa ChartGo

<https://www.chartgo.com/home.html>

ChartGo to internetowy program do tworzenia wykresów. Pozwala tworzyć bogate i kolorowe wykresy. Aby rozpocząć obrazowanie danych użytkownik powinien kliknąć na przycisk "Create Chart" (stwórz wykres). Ewentualnie użytkownik może wybrać wykres z poniżej znajdującej się karuzeli, aby użyć go jako szablonu. Ponadto, strona oferuje przewodnik i stronę Video Demo, gdzie można uzyskać więcej informacji na temat oferowanych przez aplikację internetową możliwości.

Niewątpliwą zaletą strony jest możliwość samodzielnego tworzenia i edytowania wykresów. Jest to funkcjonalność którą warto wykorzystać w projekcie systemu rankingu stron. Dzięki niej wyszkolony użytkownik systemu (pracownik firmy "Optymalizator") ma szansę w szybki, atrakcyjny i prosty sposób zwizualizować najważniejsze dla klientów dane dotyczące rankingu ich strony internetowej.

Drugą interesującą funkcją jest przewodnik. Jednakże jego formuła w systemie rankingu byłaby przeddefiniowana. Biorąc po uwagę, że klient firmy "Optymalizator" ma mieć wgląd w zaawansowane raporty zawierające informacje o pozycji jego strony w czasie i pozycji stron konkurencyjnych dla zadanej frazy, system powinien mieć przyjazny interfejs o intuicyjnym sposobie prezentacji danych. Przewodnik zatem, zamiast opisywać sposób tworzenia wykresów, byłby wsparciem dla klientów firmy "Optymalizator" chcących sprawnie poruszać się wśród zaprezentowanych zestawów wykresów i danych.

Projekt nie przewiduje nagrywania filmów instruktażowych, dlatego też sekcja Video Demo nie znajdzie zastosowania w tworzonym systemie. Ze względu na to, że firma "Optymalizator" nie prowadzi aktywnej kampanii w mediach społecznościowych, przyciski w sekcji "Follow Us" (śledź nas) przekierowująca do Facebooka, Twittera i Youtube nie znajdują się w nowo tworzonym systemie.

3. Wnioski do budowanego systemu

Budowany system składa się z dwóch podstawowych części:

- A. Części backendowej realizującej zadania związane z przetwarzaniem danych
- B. Części frontednowej podzielonej na dwa działy:
 - a. dział dla pracowników firmy "Optymalizator"
 - b. dział dla klientów firmy "Optymalizator"

Część backendowa (A) odpowiedzialna jest na regularne zbieranie, przechowywanie i dostarczanie pakietów danych powiązanych z danym klientem. Na wzór robotów indeksujących Google powinna zbierać dane w regularnych odstępach czasowych wyznaczonych przez pracownika firmy. Ponadto, dzięki automatyzacji zadań, algorytm wyszukujący powinien mieć umiejętność korzystania z już zgromadzonych danych aby zaoszczędzić zasoby obliczeniowe sprzętu.

Część frontedowa (B) realizuje dwa zadania: tworzenie wykresów dzięki danym uzyskanym z backendu oraz serwowanie przetworzonych danych (wykresów) w części aplikacji wydzielonej dla klientów firmy "Optymalizator". Dodatkowym elementem w panelu klienta jest dział "Przewodnik" który ma za zadanie wyjaśnić sposób odczytu danych z wykresów.

4. Definicja funkcji budowanego systemu

1. Systemu do tworzenia rankingu stron
2. Logowanie/Rejestracja
3. [Pracownik] Uruchomienie zbierania danych o rankingu stron
4. [Pracownik] Tworzenie wykresów
5. [Klient] Wyświetlanie wykresów i przegląd danych

