

Łukasz Niewiński,

Rekrutacja Semantive, zadanie rekrutacyjne 14.02-21.02.2020

Komentarz rozwiązania:

Aplikacja w całości oparta na języku Python, składa się z kontenerów(web, img_retrieve, textretrieve, database). Aplikacja jest dostępna dla użytkownika przeglądarkowego oraz bezpośrednio poprzez API. Obecna funkcjonalność ogranicza się do pobierania oraz przechowywania tekstu pobranego ze stron internetowych(tekst pobierany za pomocą biblioteki html2text). Nie udało mi się jeszcze rozwinąć funkcjonalności pobierania oraz przechowywania obrazów, chcę to zrobić za pomocą BeautifulSoup lub Scrapiego. Obecnie pobieram obrazy z pomocą identyfikowania tagów poprzez wykorzystanie bs4, napotkałem jednak duże problemy z przesyłaniem plików oraz komunikacją pomiędzy kontenerami. Obrazy postanowiłem przechowywać jako pliki a w bazie danych zachować jedynie ścieżki plików – te rozwiązanie wydaje się być szybsze niż zapisywanie obrazów do dodatkowej bazy danych.

Mam sporo pomysłów, jak można polepszyć/poszerzyć projekt, poza ukończeniem pozostałych funkcjonalności(img_retrieve, status check, tests), planuję również dodać gunicorna, nginxa oraz transakcje(zapisywanie danych w db).

Funkcjonalności: pobieranie oraz przechowywanie obrazów, sprawdzanie statusu i uruchomienie testów nie zostały jeszcze rozwinięte.

Podsumowanie:

Czerpałem dużo radości z pracy przy projekcie, sporo też się nauczyłem. Z chęcią przyjmę uwagi co do słabych stron proponowanego przeze mnie rozwiązania.