

Hakaton

Wyszukiwarka ofert wynajmu i kupna mieszkań

Raport 1

Repozytorium na GitHubie: <https://github.com/lpawluczuk/FlatSearch>

System zarządzania taskami Trello:

<https://trello.com/b/SSe3NyvP/flat-searcher>

Co zostało zrobione

Do tej pory udało nam się:

- założyć repozytorium oraz system zarządzania taskami w ramach Trello,
- uruchomić i skonfigurować zdalny serwer,
- wstępnie skonfigurować oraz uruchomić Solra na zdalnym serwerze,
- napisać skrypt w Pythonie crawlujący stronę Gumtree; skrypt wykorzystuje framework Scrapy,
- scrawlować Gumtree - pobrano 37 736 ogłoszeń, w sumie 46.7MB danych

Co zostanie zrobione w przyszłym tygodniu

W kolejnym tygodniu planujemy:

- “nakarmić” Solra scrawlowanymi ogłoszeniami,
- dodać kolejne serwisy do crawlera,
- stworzyć wstępną wersję front-endu.

Co ostatecznie zostanie zrobione

Nasz finalny produkt będzie wyszukiwarką ofert wynajmu i kupna mieszkań o następującej specyfikacji:

- baza scrawlowanych ogłoszeń zostanie utworzona jednorazowo (nie będziemy jej aktualizować),

- baza będzie zawierać co najmniej 100 000 ogłoszeń,
- ogłoszenia będą pochodzić z serwisów Gumtree, Tablica, OtoDom, Gratka, Regiodom, Morizon,
- baza będzie obejmowała oferty z całej Polski,
- z ogłoszeń zostaną wyciągnięte (również za pomocą wyrażeń regularnych) następujące parametry: tytuł, treść ogłoszenia, data dodania, cena, liczba pokoi, lokalizacja, metraż, rodzaj zabudowy,
- wyszukiwarka będzie umożliwiała filtrowanie wyników w ramach wyżej wymienionych parametrów.

Za pomocą czego wykonamy zadanie

Z “technicznego” punktu widzenia:

- na zdalnym serwerze zostanie postawiona i skonfigurowana wyszukiwarka Solr,
- za pomocą Pythona i frameworka Scrapy zostaną wyciągnięte ogłoszenia z wymienionych wcześniej serwisów,
- JSONy, które w efekcie otrzymamy, zostaną poddane obróbce (również za pomocą Pythona), aby można “nakarmić” nimi Solra,
- przy użyciu technologii HTML/CSS/JavaScript stworzony zostanie front-end komunikujący się z Solrem (prawdopodobnie wykorzystamy bibliotekę SolrJS).