WARMONGER PROJECT: CRAWLER.

Парсер для базы данных

MozhaykaTeam

Проблематика

Насыщение баз данных, содержащих информацию о комплектующих БПЛА в настоящее время осуществляется в ручном режиме.

Различная структура web-страниц для парсинга

Отсутствие универсального средства сбора информации - метапауков

Подход к решению задачи

1. Автоматизированный сбор информации (Краулеры, основанные на структуре сайта)
2. Автоматический сбор информации (Метапаук)
3. Использование языковой модели для вычисления семантической близости найденных сущностей
4. Визуализация собранных данных
5. Оценка дальности полета БПЛА

Стек технологий

VueJS

YandexMapAPI

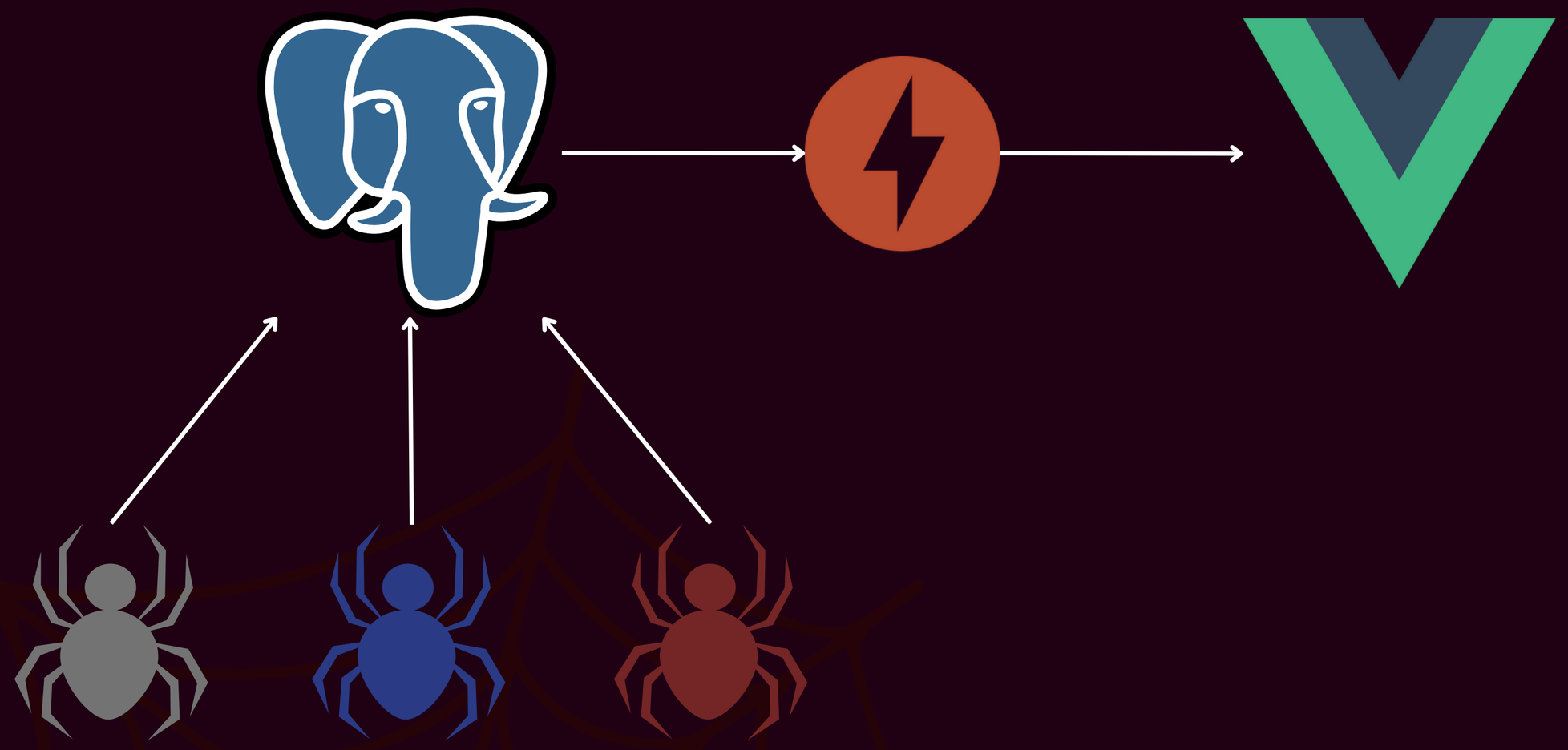
Scrapy

PostgreSQL

PyTorch

FastAPI

Архитектура решения



Технические особенности

1 Краулеры вынесены за пределы FastAPI в целях безопасности (secure by design)

2 Использована гипотеза о том, что целевая информация хранится в определенного типа html-тегах, что позволяет построить такой краулер, который находил бы эту информацию вне зависимости от структуры сайта

3 Оценка дальности полета предназначена для обзора БПЛА различных государтсв с возможностью фильтрации по странам, а также нанесения метки выбранного БПЛА на любую локацию карты мира для оценки дальности полета для симулирования возможностей БПЛА.

Ценность проекта

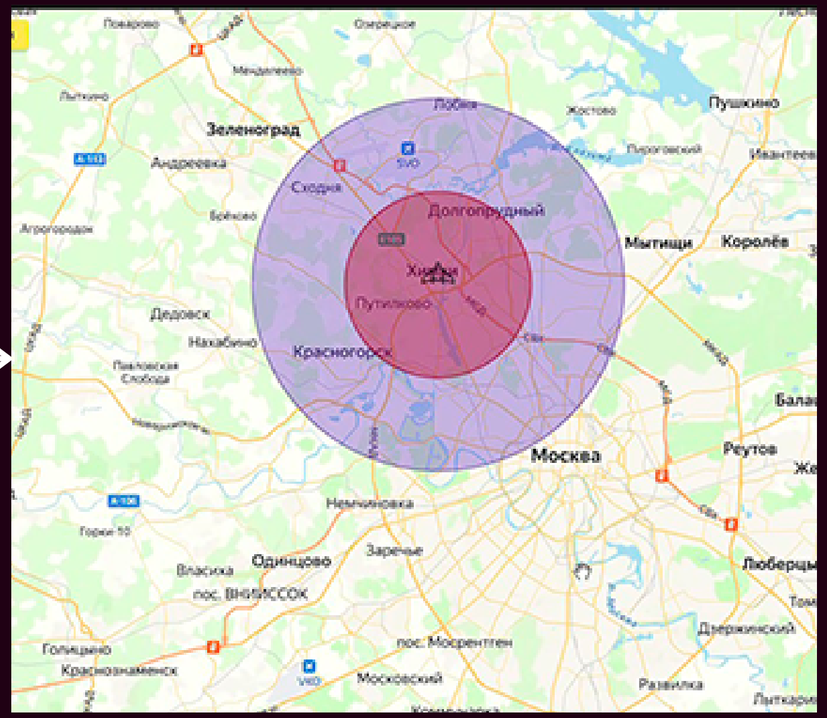
Парсинг любого сайта

Scrapy показывает высокие скоростные параметры

Визуализация и симуляция использования объектов

Высокая адаптируемость и масштабируемость

Оценка дальности полета



Перспективы разработки

Повышение точности

Добавление новых вкладок для визуализации

Добавление новых типов структур для метапаука

Исследование новых методов вычисления семантической близости

Команда проекта

Артем
Team-lead
@qTemio

Георгий
Data Science
@Focus_Fire

Игорь
Front-end
@YourCarma1

Захар
Front-end
@lildubber