

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«ПЕТРОЗАВОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Институт математики и информационных технологий

Отчет о прохождении производственной практики

Выполнила студентка группы 22307:
Светова Кристина Сергеевна

Направление подготовки:
09.03.04. Программная инженерия

Место прохождения практики:
“HERE Technologies (ООО «ГЕА РУ
Си-АйЭс»)”

Сроки прохождения практики:
30.05.2022 - 09.06.2022

Руководитель практики:
к.т.н, доцент
Богоявленская Ольга Юрьевна

Оценка _____

Дата _____

Петрозаводск 2022 г.

О компании

Технологическая компания “HERE Technologies” создана в 1985 году и имеет 30 летнюю историю. Офисы компании расположены в 55 странах мира, везде, где люди активно пользуются автомобилями.

На данный момент, HERE Technologies принадлежит немецкому консорциуму автомобильных производителей: BMW, Audi и Daimler (Mercedes-Benz).

HERE Technologies – является глобальным и крупнейшим поставщиком навигационных карт для производителей автомобильной техники по всему миру.

Компания поддерживает не только свои традиционные продукты – навигационные карты, но и активно принимает участие в разработке инновационных проектов, связанных с роботизированным вождением, искусственным интеллектом, большими данными, уникальными сервисами.

Ход практики

Вводная презентация

В первый день практики была организована вводная презентация с представителями компании для знакомства, понимания целей практики и того, чем предстоит заняться.

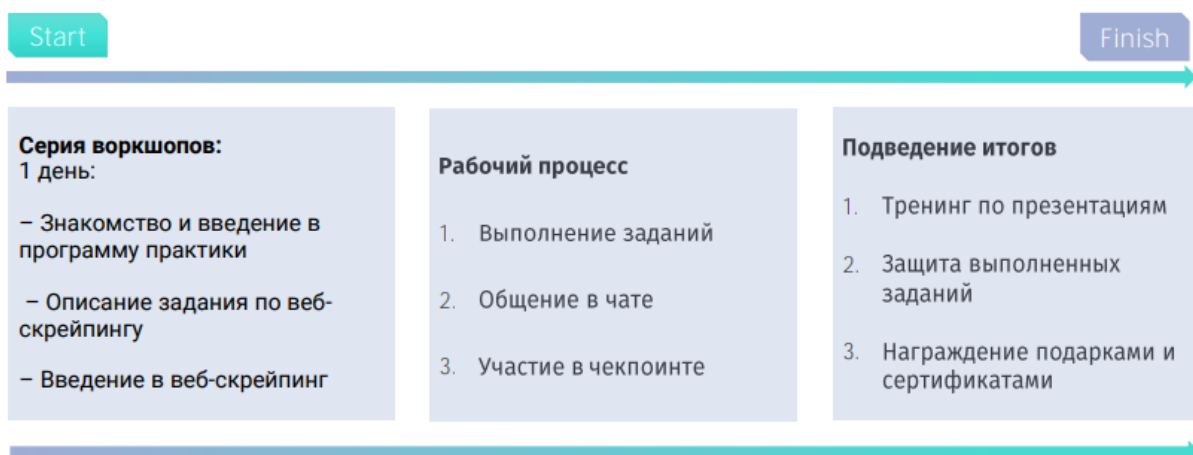
Направление практики - Data Engineering: сбор, анализ и визуализация пространственных данных.

Были выделены следующие задачи на практику:

- ознакомиться с обучающими материалами
- освоить сервисы HERE Platform, Python, Scrapy
- разработать модели для парсинга сайтов
- подготовить отчетные материалы, опубликовать данные на GitLab

Были также установлены этапы, которые мы будем проходить, это:

- серия воркшопов с обучающими материалами для понимания основ при выполнении ключевых задач
- рабочий процесс, включающий участие в чекпойнтах для получения связи с кураторами о твоём продвижении по задачам, возможности обсудить появившиеся проблемы и их решения
- защита презентаций об итогах прохождения практики



Поставленные задачи

После ознакомления с вводной частью приступили к получению практических знаний в сфере веб-скрейпинга. Важной частью этого этапа было изучение обучающих материалов и посещение воркшопов.

Для выполнения заданий использовали инструменты командной разработки такие как: Trello - для отслеживания прогресса и постановки заданий участникам, Gitlab - для публикации своих результатов.

В работу требовалось взять не менее 2х сайтов для веб-скрейпинга, мой выбор был следующим:

- Gyongy Patika (Венгрия)
- Evpoint (Болгария)
- Рерсо (Сербия)
- Aldi (Венгрия)

Ход выполнения

В первую очередь выбранные сайты нужно было проанализировать: понять архитектуру построения веб-сервиса, познакомиться с пользовательскими соглашениями на момент разрешения сборки данных с сервиса, изучение способа выгрузки пространственных данных.

После выполнения аналитической части нужно было приступить к реализации модели парсинга данных. Данная часть выполнялась с помощью следующего стека технологий: VS code, Google Collaboratory, Python, Scrapy, Pandas, BeautifulSoup, Selenium, Requests, Gitlab.

В результате полученная модель должна была представлять полученные пространственные данные в виде файла GeoJson для возможности последующего анализа и визуализации данных.

Дополнительным этапом удалось поработать с геокодерами, попробовать оба варианта: прямое и обратное геокодирование для анализа проверки качества собранных данных. Для реализации этой части использовался фреймворк Scrapy с Geocoder API.

Подведение итогов

По окончании была проведена защита наших результатов. Была подготовлена презентация с описанием полученных результатов и визуализацией данных для выбранных четырех слайдов.

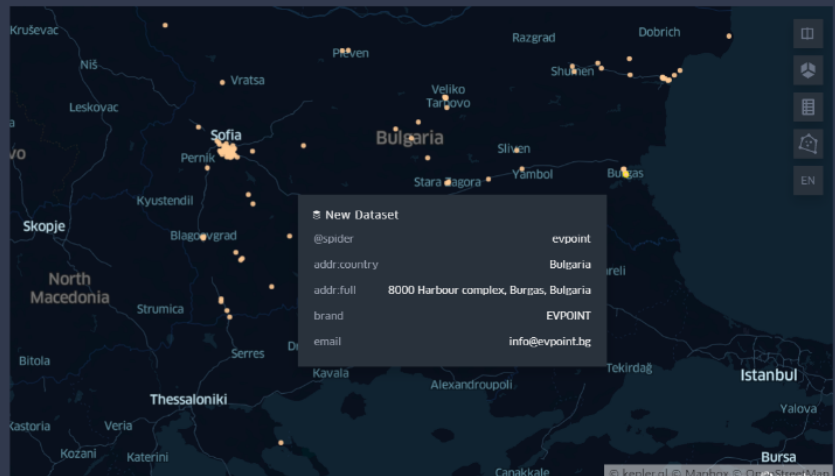
Получение данных

Evpoint : 144 точки

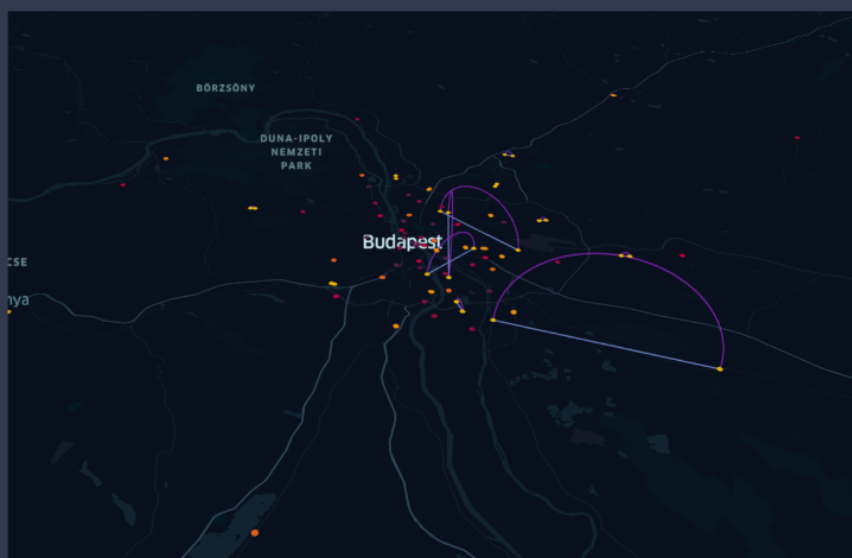
Get - запрос

Получены:

- Адрес
- Название точки
- Телефон
- Эл.адрес
- координаты



ALDI



	distance	query_score
count	92.000000	92.000000
mean	905.368302	0.950326
std	3403.656848	0.083317
min	1.006061	0.610000
25%	27.067792	0.927500
50%	58.207541	1.000000
75%	245.312250	1.000000
max	24738.637652	1.000000

В результате за время практики удалось изучить основы области веб-скрейпинга, получить практические знания по разработке моделей извлечения пространственных данных с помощью фреймворка Scrapy, а также проанализировать и визуализировать данные, получить опыт работы с геокодерами. Огромным плюсом было использование командных средств разработки и постоянное общение с кураторами для получения обратной связи по каждому из этапов.