



HERE Data Engineering

HERE Technologies

2022

HERE Technologies



образованы в
1985



7000
сотрудников



охватываем 5
континентов



56 стран
с офисами HERE



Наши партнёры

Сегодня

№1 платформа данных
и геолокационных
технологий

#1



HERE лидер отрасли

По данным Omdia, Counterpoint
Research и Strategy Analytics

1-е место по полноте
в отрасли

Превосходя Google, Mapbox,
TomTom и Apple

Развитие активных сообществ

С 2012 компания активно сотрудничает с университетами по всему миру, создавая совместные проекты, мероприятия и обучающие программы.



Наши приоритеты:



Развитие активного сообщества



Улучшение профессиональных навыков



Постоянная экспертная поддержка



Web Scraping

Зачем нужен проект?



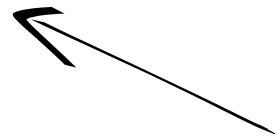
Эффективность

Сбор данных на основе построения web scraping моделей (скриптов).



IT КОМЬЮНИТИ

Работа в сообществе.
Возможность обмена опытом в международном сегменте.



Развитие

Возможность роста, участие в проекте в качестве амбассадора и модератора.
Выход на фриланс работу.



Эффективность

Почему это выгодно?



Большие масштабы



Высокое качество



Автоматический процесс



Применимо для любой сферы деятельности





IT КОМЬЮНИТИ

Нетворкинг



Личный бренд



Поддержка



Новые возможности





Развитие



Программа
Чего ждать?

HERE Data Engineering

онлайн программа по сбору, анализу и визуализации пространственных данных

Задачи:

- ✓ Освоить сервисы [HERE Platform](#), Python, Scrapy
- ✓ Разработать модели для парсинга сайтов
- ✓ Подготовить отчётные материалы, опубликовать данные на GitLab



Какие навыки освоите за период программы?



Описание этапов

Start

Finish

Серия воркшопов

- Знакомство и введение в программу практики
- Описание задания по веб-скрейпингу
- Введение в веб-скрейпинг

Рабочий процесс

1. Выполнение заданий
2. Общение в чате
3. Участие в чекпоинте

Подведение итогов

1. Тренинг по презентациям
2. Защита работ
3. Награждение подарками и сертификатами

Расписание (пример)

Пн	Вт	Ср	Чт	Пт	Сб	Вс
24 <u>16:30</u> Знакомство и введение	25	26 <u>16:00</u> Web Scraping Часть 1	27	28	29	30
31 <u>16:30</u> Web Scraping Разбор задания	1	2 <u>16:00</u> Чекпоинт	3	4	5	6
7 <u>16:30</u> Чекпоинт	8	9 <u>16:00</u> Чекпоинт	10	11	12	13
14 <u>16:30</u> Web Scraping Визуализация	15	16 <u>16:00</u> Чекпоинт	17	18	19	20
21 <u>16:30</u> Мастер класс по презентациям	22	23 <u>16:00</u> Чекпоинт	24	25	26	27
28	29	30 <u>16:00</u> Защита работ, выступления	1	2	3	4

Октябрь

Ноябрь

***Чекпоинты** – онлайн встречи для обсуждения статуса задания и разбора сложностей и вопросов

Applications



Trello board

Пространство для трекинга сайтов



Telegram

Чат для общения, вопросов



Webex meeting

Для звонков



Git Lab

Для управления кодовой базой



Python

Для программирования



Visual Code

Среда разработки

Техническая сторона проекта

Примеры веб-сайтов для анализа

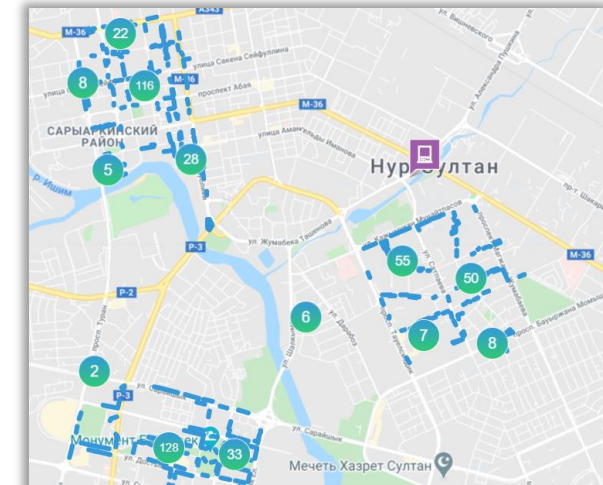
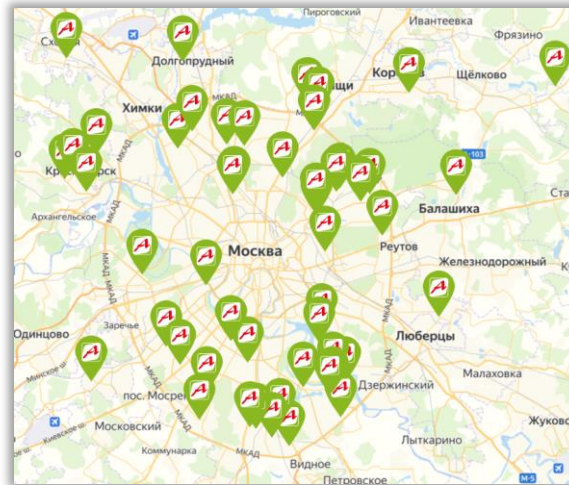
Информационные ресурсы для скрейпинга:

- [Сеть магазинов “Авоська” - avoska.ru/](https://avoska.ru/)
- [Магнит Косметик - magnitcosmetic.ru/shops/](https://magnitcosmetic.ru/shops/)
- [Билайн - beeline-tochki.ru/store](https://beeline-tochki.ru/store)

Требуемые данные:

- Адреса точек интереса желательно с геопривязкой (координаты);
- Контактная информация;
- Часы работы;
- Другая сопутствующая информация.

г. Москва, ул. 1812 года, 12	08:00 - 23:00
г. Москва, ул. Абельмановская, 6	10:00 - 21:00
г. Москва, ул. Адмирала Лазарева, 40, помещение IV	11:00 - 20:00
г. Москва, ул. Азовская, 24,2,этаж.1.пом.II.ком.1-8.1718.20-26	10:00 - 21:30
г. Москва, ул. Академика Скрябина, 14, корп 1, этаж 1	09:00 - 21:00
г. Москва, ул. Алма-Атинская, 7	10:00 - 21:30
г. Москва, ул. Бакунинская, 69, стр. 1, помещения №V комн.1, пом. VI комн. 1-4	10:00 - 22:00
г. Москва, проезд Батайский, 35, 1 этаж, номера комнат на поэтажном плане: 1, 4-9, 11, 12, 15, 16	10:00 - 22:00
г. Москва, ул. Беловежская, 63	10:00 - 22:00
г. Москва, ул. Бориса Галушкина, 20, этаж 1	10:00 - 22:00
г. Москва, ул. Братиславская, 26, комнаты 1, 2, 3, 3а, 5, 6, 7, 8, 8а, 8б, 8в, 8г, 8д, 8е, 9, 9а, 9б, 9в, 10б, 10г, 10д, 10е, 10ж, 10з, 10и, 10к, 10л	10:00 - 21:30
г. Москва, ш. Варшавское, 128, корп. 2, подвал, помещение III - комнаты 1,2; этаж 1, помещение I - комнаты с 53 по 77	09:00 - 21:00
г. Москва, ш. Варшавское, 141, корп. 9А	10:00 - 22:00
г. Москва, ул. Верхние Поля, 45, корп.3	10:00 - 22:00



Перечень рекомендуемых технологий



Обработка данных на Python:

- [Scrapy](#) (скрейпинг)
- [BeautifulSoup](#) (библиотека для парсинга)
- [Requests](#) (HTTP запросы)
- [Pandas](#) (обработка данных)
- [GeoPandas](#) (обработка пространственных данных) [IPyleaflet](#) (надстройка для Jupyter Notebook, для отображения данных на карте)

Визуализация результатов:

- [HERE Maps API for JavaScript](#)
- [HERE Studio & Data Hub](#)
- [HERE Geojson Viewer](#)

Подведение итогов

Награждение

1

3 сайта – минимальное количество для защиты

2

По итогу помогаем с отчётами, предоставляем характеристики, дневники практики по запросу

3

Подарки в конце программы – тип подарка на усмотрение сотрудников HERE

Критерии оценки итоговых работ

- ✓ Уровень сложности реализации;
- ✓ Количество распарсенных сайтов;
- ✓ Качество данных;
- ✓ Качество защиты.



Контакты



Максим Макаров



@maxmakarov



maksim.makarov@here.com

Виктория Ермакова



@VErmakova



viktoria.ermakova@here.com